



CRIME PREDICTION AND HOTSPOT MAPPING

A MINI PROJECT REPORT

Submitted by

SINDHUJA S(1920110047)

KARTHI SARAVANA O(1920110702)

PRAVEENRAJ R(1920110704)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

SONA COLLEGE OF TECHNOLOGY, SALEM-5

(Autonomous)

ANNA UNIVERSITY: CHENNAI 600 025

November 2023

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report “**Crime Prediction and Hotspot Mapping**” is the bonafide work of “**SINDHUJA S(1920110047), PRAVEENRAJ R(1920110704), KARTHI SARAVANA O(1920110702)**” who carried out the project work under my supervision

SIGNATURE

Dr. J. Akilandeswari

HEAD OF THE DEPARTMENT

Professor

Department of Information Technology

Sona College of Technology

Salem- 636 005.

SIGNATURE

Ms.Sangeetha priya

SUPERVISOR

Assistant Professor

Department of Information Technology

Sona College of Technology

Salem- 636 005.

Submitted for Project viva voce examination held on _____

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

First and foremost, we thank to **power of almighty** for showing us inner peace and for all blessings .Special gratitude to our parents, for showing their support and love always.

We express our sincere thanks to chairman, **Sri.C.VALLIAPPA** and principal **Dr.S.R.R.SENTHILKUMAR** for providing adequate facilities to complete the project.

We are immensely grateful to our Head of Information Technology Department, **Dr.J.AKILANDESWARI** for the continuous encouragement to complete the project.

We express our heartfelt thanks to our project supervisor **Ms.SANGEETHA PRIYA** for her valuable guidance and fruitful discussions throughout the course of the project work.

We feel proud in sharing this success with all our department Faculty, Staff members and friends who helped directly and indirectly in completing this project successfully.

ABSTRACT

Crime is a major social and economic problem in almost every country, which threatens the safety of its citizens and also disrupts the economy of that nation. Understanding patterns in criminal activity will allow us to predict the crimes that may occur in the future and also predict their **“hotspots”** (the areas where they are most prominent to occur) and enables the authorities to more effectively and efficiently allocate manforce and resources to prevent or respond to incidents.

This Streamlit web application leverages Python, Streamlit, and various libraries for real-time incident reporting, crime data visualization, and major possible crime occurrence area prediction. The backend is powered by Pandas for data manipulation and analysis, Geopandas for geographic data handling, and Folium for interactive mapping. Users can report incidents through an intuitive Streamlit interface, with incident data dynamically stored in a CSV file for ongoing analysis. The application features a map-based visualization of historical crime data, including hotspots represented through a heatmap. To enhance user awareness, a K-Means clustering algorithm is applied to predict major possible crime occurrence areas, providing an additional layer of insight. The application aims to empower users and law enforcement with actionable insights derived from geospatial crime data. This abstract provides more details about the technologies used, including Python, Streamlit, Pandas, Geopandas, and Folium. It emphasizes the seamless integration of these technologies to create a user-friendly and informative crime prediction and reporting web application.

LIST OF FIGURES

Fig.1.1	Example Figure	11
Fig.1.2	Sample Data	16
Fig.1.3	Sample Analysis	16
Fig.2.1	K-means Diagram	20
Fig.2.2	Dataset	24
Fig.2.3	Flow Chart	26
Fig.2.4	Architecture of Data processing	28
Fig.2.5	Crime Predict and Mapping	31
Fig.2.6	Accuracy and Average	31
Fig.2.7	Report an Incident	32
Fig.2.8	Major Possible Crime	32
Fig.2.9	Location Data Table	33

LIST OF ABBREVIATIONS

PCA	- Principal Component Analysis
GIS	- Geographic Information Systems
API	- Application Programming Interface
EDA	- Exploratory Data Analysis
LNP	- Natural Language Processing
RNN	- Recurrent Neural Network
STNN	- Spatio Temporal Neural Network

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	iv
	LIST OF FIGURES	v
	LIST OF ABBREVIATIONS	vi
1.	INTRODUCTION	8
	1.1. Problem Statement	9
	1.2. Objective	9
	1.3. Crime Prediction Overview	10
	1.4. Literature survey	13
	1.5. Existing System	15
	1.6. Disadvantages	16
2.	SYSTEM ARCHITECTURE	18
	2.1. Technology used	18
	2.2. Implementation Details	25
	2.3. Modules	28
	2.4. User Interface Design	29
3.	CONCLUSION AND FUTURE WORKS	34
	REFERENCES	36
	APPENDIX - I	37

CHAPTER 1

INTRODUCTION

Crime prediction and hotspot mapping are cutting-edge tools that law enforcement agencies employ to enhance public safety by using data-driven approaches. These techniques harness the power of advanced machine learning and spatial analysis to predict when and where crimes are likely to occur, enabling law enforcement to allocate resources more effectively. Crime prediction involves the utilization of historical crime data and machine learning algorithms to identify patterns and anticipate future criminal activities. Simultaneously, hotspot mapping employs geographic information systems and statistical methods to pinpoint areas with significantly higher crime rates, thus enabling law enforcement to focus efforts on these hotspots.

These methodologies enable law enforcement to take proactive measures, optimize resource allocation, and enhance community safety through a data-centric approach, while also addressing ethical and privacy considerations in their implementation.

Crime prediction and hotspot mapping have emerged as essential components of modern law enforcement and public safety strategies.

These methodologies are rooted in data analysis, allowing law enforcement agencies to move from a reactive to a proactive stance in addressing criminal activities. By analyzing historical crime data, machine learning algorithms can uncover intricate patterns and relationships within the data, which can be used to forecast potential future criminal incidents. This predictive capacity enables law enforcement to anticipate where and when crimes might happen, facilitating more strategic deployment of resources.

These data-driven techniques, however, come with a set of ethical considerations that must be addressed. Ensuring fairness and impartiality in predictive models, as well as respecting individual privacy rights, is crucial in their implementation.

Ethical considerations are paramount to maintaining public trust and preventing discrimination or bias in law enforcement practices.

crime prediction and hotspot mapping represent a significant shift in law enforcement strategies. They empower law enforcement agencies with the tools and insights needed to proactively combat crime, allocate resources efficiently, and improve community safety, all while upholding ethical and privacy standards. These methodologies are a testament to the power of data-driven decision-making in shaping the future of public safety efforts.

1.1 PROBLEM STATEMENT

Crime prediction and hotspot mapping are critical tasks in law enforcement and public safety. The goal is to develop a data-driven system that can analyze historical crime data, identify patterns, and predict where future crimes are likely to occur. This system should also provide law enforcement agencies with valuable insights to allocate resources effectively and proactively prevent crimes in specific areas.

1.2 OBJECTIVES

- **Crime Prediction:** To accurately predict where and when future crimes are likely to occur, enabling law enforcement agencies to take proactive measures to prevent criminal activities.
- **Hotspot Mapping:** To identify areas with a high concentration of past and potential future crimes, allowing law enforcement to focus their resources on these hotspots.
- **Resource Allocation:** To provide law enforcement agencies with valuable insights on where to allocate personnel and resources most effectively, based on predicted crime trends.
- **Crime Reduction:** To contribute to the reduction of crime rates in specific areas by targeting preventive measures and law enforcement presence in high-risk zones.

- **Public Safety:** To enhance public safety and community well-being by improving law enforcement's ability to respond to and prevent criminal incidents.
- **Efficiency:** To optimize the allocation of limited law enforcement resources, leading to more efficient use of personnel, time, and budget.
- **Data-Driven Decision Making:** To promote evidence-based decision-making in law enforcement by leveraging historical data and predictive analytics.
- **Transparency:** To ensure transparency in the decision-making process, enabling law enforcement agencies and the public to understand how resource allocation decisions are made.
- **Continuous Improvement:** To establish a system that can adapt and improve over time as new data and insights become available, thereby increasing its accuracy and effectiveness.
- **Ethical Considerations:** To address potential ethical and privacy concerns, ensuring that the system respects individuals' rights and complies with relevant regulations and laws.

1.3. CRIME PREDICTION OVERVIEW

Machine learning and Artificial Intelligence have gained immense popularity, starting from the end of the previous decade. Their share, almost unlimited scope and areas of application has taken over the decade by storm. To humans research and analytics for predicting the chance of some event occurring has always been one of the top priority. Such type of analysis is broadly classified as predictive analysis. This predictive analysis helps us gain a possible insight in the future. This insight in predicting storms, natural calamities, crimes, etc can mean the difference between losing and saving millions.

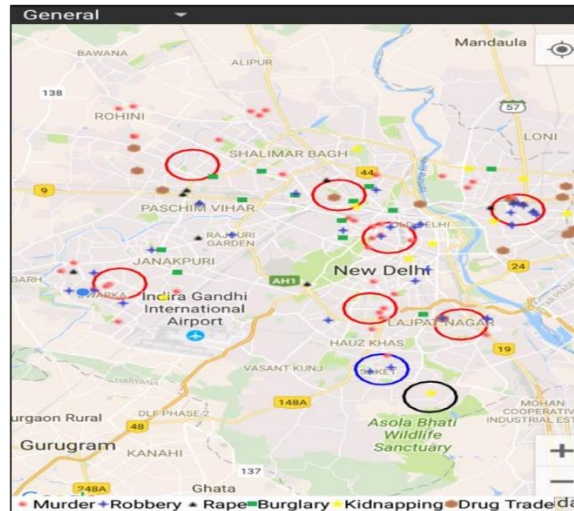


Fig.1.1 Examples

1.3.2. Challenges and Considerations

When implementing a crime prediction system, several challenges and considerations should be taken into account to ensure the system's effectiveness, ethical compliance, and public trust. Here are some key challenges and considerations:

Data Quality and Availability:

- **Data Quality:** Crime prediction relies on historical crime data. Ensuring the accuracy, completeness, and consistency of this data is crucial. Inaccurate or incomplete data can lead to misleading predictions.
- **Data Privacy:** Balancing the need for data with privacy concerns is a significant challenge. Collecting and using personal information must adhere to legal and ethical standards to protect individuals' rights.
- **Data Bias:** Historical crime data may be biased due to various factors, including underreporting, police practices, or systemic biases. These biases can affect the reliability of predictions and resource allocation decisions.

Resource Allocation:

- **Resource Allocation Optimization:** Allocating resources effectively and efficiently based on predictions is a complex challenge. Law enforcement agencies need to balance budget constraints and personnel deployment.
- **Evaluating Impact:** It's essential to continually evaluate whether the allocation of resources based on predictions is having the intended impact on reducing crime rates.

Model Development and Evaluation:

- **Overfitting:** Preventing overfitting is a challenge when using complex predictive models. Proper validation techniques and hyperparameter tuning are necessary to ensure the model generalizes well.
- **Model Transparency:** Ensuring that predictive models are interpretable and explainable is important for gaining public trust and making informed resource allocation decisions.
- **Temporal Patterns:** Capturing and understanding temporal patterns in crime data, such as seasonality or trends, is essential for accurate predictions.

Training and Expertise:

- **Training Law Enforcement:** Ensuring that law enforcement personnel are adequately trained to use predictive models and understand their limitations is essential for effective implementation.

Continuous Improvement:

- **Model Updates:** Crime prediction models should be updated regularly to incorporate new data and adapt to changing crime patterns.
- **Feedback Loops:** Establish feedback mechanisms that allow law enforcement agencies to provide insights and expertise to data scientists for model improvement.

1.4. LITERATURE SURVEY

- **Chainey, S., & Ratcliffe, J. (2005). GIS and Crime Mapping. John Wiley & Sons.**

This book provides an in-depth introduction to the use of Geographic Information Systems (GIS) in crime mapping and spatial analysis. "GIS and Crime Mapping" by Chainey and Ratcliffe, published in 2005, is an in-depth introduction to the use of Geographic Information Systems (GIS) in the context of crime mapping and spatial analysis. This comprehensive resource covers GIS fundamentals, crime mapping techniques, spatial analysis, data sources, practical applications, and crime prevention strategies, making it an essential guide for understanding how GIS technology can enhance crime analysis and public safety efforts.

- **Mohler, G., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. Journal of the American Statistical Association, 106(493), 100-108.**

This paper discusses self-exciting point process models and their application in crime prediction. The 2011 article "Self-exciting point process modeling of crime" by Mohler, Short, Brantingham, Schoenberg, and Tita, published in the Journal of the American Statistical Association, presents a statistical approach that utilizes self-exciting point processes to model crime incidents. This research explores the dynamic and self-propagating nature of criminal activity, aiming to improve predictive and analytical capabilities in the field of crime analysis by accounting for spatial and temporal dependencies among crime events.

- **Ashby, M. P. J., & Bowers, K. J. (2012). Prospective Hot-Spotting: The Future of Crime Mapping? British Journal of Criminology, 52(2), 381-398.**

This article explores the concept of prospective hotspot mapping and its potential impact on proactive policing. In the 2012 article "Prospective Hot-Spotting: The Future of Crime Mapping?" by Ashby and Bowers, published in the British Journal of Criminology, the authors discuss the concept of prospective hot-spotting, which represents a shift in crime mapping from reactive analysis to proactive prediction.

This innovative approach leverages historical crime data and other contextual information to forecast future crime hotspots, providing law enforcement with a valuable tool for crime prevention and resource allocation, thus offering insights into the evolving landscape of crime analysis and prediction.

- **Johnson, S. D., Bernasco, W., Bowers, K. J., Elffers, H., Ratcliffe, J., Rengert, G. F., & Townsley, M. (2007). Space–time patterns of risk: A cross national assessment of residential burglary. Journal of Quantitative Criminology, 23(3), 201-219.**

This study investigates space-time patterns of residential burglary, which is a key aspect of hotspot mapping. The 2007 study "Space–time patterns of risk: A cross-national assessment of residential burglary" by Johnson, Bernasco, Bowers, Elffers, Ratcliffe, Rengert, and Townsley, published in the Journal of Quantitative Criminology, conducts a comprehensive cross-national analysis of residential burglary patterns, examining the spatial and temporal aspects of risk. This research provides valuable insights into the dynamics of residential burglary across different regions, contributing to a better understanding of crime and the development of effective crime prevention strategies and policies.

- **Mohler, G., Short, M. B., & Malinowski, S. (2016). Random forests and the prediction of burglary. Journal of the American Statistical Association, 111(513), 739-748.**

This paper delves into the application of random forests in predicting burglaries. The 2016 paper "Random Forests and the Prediction of Burglary" by Mohler, Short, and Malinowski, published in the Journal of the American Statistical Association, explores the use of Random Forests, a machine learning technique, for the prediction of burglary incidents. This research demonstrates the effectiveness of Random Forests in forecasting burglary events, offering a data-driven and predictive approach that has practical applications in crime prevention and law enforcement by enhancing the accuracy of crime predictions.

1.5. Existing System

The existing system for crime prediction and hotspot mapping relies heavily on traditional crime analysis methods and the experience of law enforcement personnel. Historical crime data is manually reviewed and analyzed to identify patterns and hotspots, with decisions and resource allocation largely guided by human expertise. While this approach has been the cornerstone of law enforcement for many years, it is disadvantaged in crime prediction and mapping. The existing system for crime prediction and hotspot mapping suffers from subjectivity, limited predictive power, inefficiency, lack of real-time capabilities, and potential ethical concerns. Human analysts may introduce bias, hindering the accuracy of hotspot identification. The reliance on historical data may not capture evolving trends, and manual analysis can be time-consuming. Real-time insights and integration with GIS technology are often lacking, limiting the system's responsiveness. Ethical concerns regarding potential bias and privacy issues may also arise in this approach.

1.5.1 Examples of how machine learning is being used for crime prediction and hotspot mapping today:

- The Los Angeles Police Department is using a machine learning algorithm called PredPol to predict crime hotspots. PredPol has been shown to be effective in reducing crime rates in the areas where it has been deployed.
- The Chicago Police Department is using a machine learning algorithm called HunchLab to help detectives identify suspects in unsolved crimes. HunchLab has been shown to be effective in helping detectives to identify suspects more quickly and accurately.

The New York Police Department is using a machine learning algorithm called CompStat to track crime rates and to identify crime trends. CompStat has been shown to be effective in helping the NYPD to deploy resources more effectively and to reduce crime rates.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	date	time_of_day	crime_type	location	latitude	longitude	victim_gender	victim_age	perpetrator_gender	perpetrator_age	weapon	injury	weather	temperature	previous_activity
2	12/18/2021	21:29:18	Assault	chennai	13.1206	77.5712	Female	62	Other	18	Gun	Major	Partly Cloudy	40	Fraud
3	9/16/2018	8:36:10	Forgery	chennai	13.068	77.5754	Male	56	Male	18	None	Fatal	Partly Cloudy	10	None
4	6/2/2018	4:27:36	Embezzlement	chennai	13.0788	77.6458	Other	18	Male	20	Gun	Fatal	Partly Cloudy	5	Robbery
5	2/24/2020	10:51:00	Robbery	chennai	12.8899	77.5503	Male	25	Male	25	Gun	Minor	Clear	25	Drug Offense
6	6/2/2018	8:01:24	Forgery	chennai	13.0214	77.4496	Male	53	Other	63	Knife	None	Clear	20	Fraud
7	2/24/2018	21:00:41	Embezzlement	chennai	13.0559	77.4998	Other	36	Female	47	Gun	Fatal	Overcast	1	Robbery
8	4/13/2019	18:54:23	Vandalism	chennai	13.0814	77.5721	Other	61	Female	39	Knife	Major	Overcast	0	Burglary
9	4/16/2021	21:58:36	Forgery	chennai	13.0547	77.5186	Female	25	Female	23	Knife	Major	Clear	25	None
10	1/13/2022	3:38:23	Forgery	chennai	12.9809	77.7966	Female	58	Female	18	Gun	Minor	Clear	0	Embezzlement
11	3/5/2018	9:25:17	Burglary	chennai	13.0961	77.4986	Female	35	Other	49	Knife	None	Clear	37	Drug Offense
12	4/29/2021	6:56:17	Theft	chennai	12.9208	77.4237	Other	30	Female	28	None	Fatal	Overcast	20	Burglary
13	8/2/2022	20:16:00	Fraud	chennai	13.1532	77.6746	Female	55	Male	48	Blunt Object	Minor	Clear	7	Forgery
14	7/5/2021	20:33:33	Robbery	chennai	12.9826	77.7132	Other	62	Female	23	None	Fatal	Rain	10	Embezzlement
15	10/1/2018	18:13:30	Vandalism	chennai	13.1391	77.62	Male	27	Female	31	None	None	Clear	33	Assault
16	9/28/2021	6:36:59	Embezzlement	chennai	12.8746	77.7989	Other	37	Female	42	Gun	Major	Rain	8	Fraud
17	6/22/2020	23:05:43	Theft	chennai	12.9363	77.4805	Male	21	Other	61	Blunt Object	Major	Partly Cloudy	17	Drug Offense
18	8/13/2022	0:37:42	Vandalism	chennai	12.9822	77.5382	Male	27	Other	52	Knife	None	Clear	3	Theft
19	5/11/2019	6:28:02	Assault	chennai	13.0582	77.6439	Other	43	Other	58	Knife	Major	Partly Cloudy	15	None
20	7/10/2022	15:26:00	Embezzlement	chennai	13.1544	77.439	Male	27	Female	22	None	Major	Overcast	27	Vandalism
21	11/30/2021	22:34:24	Burglary	chennai	13.0076	77.5068	Other	62	Female	57	Knife	Major	Snow	40	Fraud
22	1/27/2021	4:25:27	Burglary	chennai	12.8912	77.4086	Male	40	Other	61	Blunt Object	Minor	Snow	30	Burglary
23	12/27/2022	15:19:12	Embezzlement	chennai	13.0014	77.4377	Male	33	Other	27	Gun	Major	Overcast	24	Theft
24	2/22/2019	23:52:51	Fraud	chennai	13.1463	77.4136	Other	25	Male	31	Gun	Major	Snow	8	Fraud
25	5/21/2018	19:11:22	Drug Offense	chennai	13.1326	77.4475	Other	46	Other	47	None	None	Partly Cloudy	11	Vandalism
26	4/14/2019	21:19:32	Assault	chennai	13.1024	77.5917	Other	18	Other	43	Knife	None	Snow	28	Fraud

Fig.1.2 Sample data

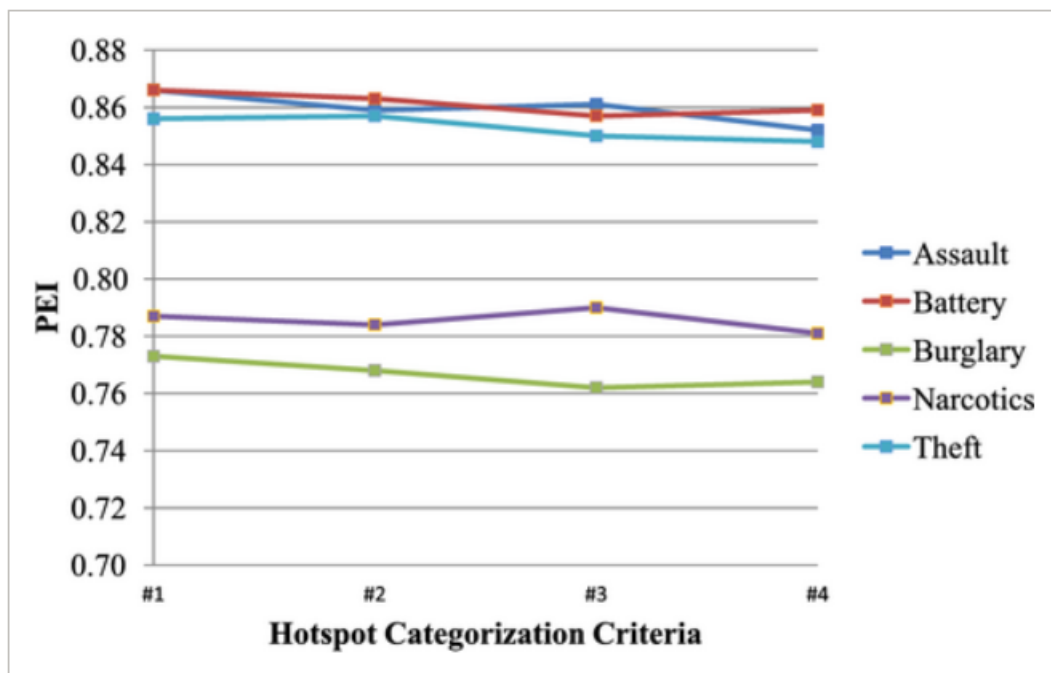


Fig.1.3 Sample Analysis

1.6. Disadvantages in Existing System

- **Subjectivity:** Human analysts may introduce subjectivity into their assessments, leading to potential bias and inconsistencies in identifying crime hotspots. This subjectivity can hinder the accuracy of resource allocation and the effectiveness of crime prevention efforts.

- **Limited Predictive Power:** The existing system often lacks sophisticated predictive models, making it challenging to anticipate future criminal activity accurately. Analysts rely on historical data alone, which may not capture evolving trends and emerging hotspots.
- **Inefficiency:** Manual data analysis is time-consuming and resource-intensive. Analysts must manually sift through large datasets, which can be an inefficient process, particularly in areas with high crime rates or complex patterns.
- **Lack of Real-Time Capabilities:** The existing system may not provide real-time insights, making it difficult to respond promptly to dynamic criminal activities or emerging hotspots.
- **Insufficient Data Integration:** Geographic Information may be used for mapping crime incidents, but the integration with predictive modeling is often limited. This can impede the ability to create accurate and up-to-date hotspot maps.
- **Ethical Challenges:** There can be ethical concerns related to potential bias or discrimination in decision-making, as well as questions about privacy when collecting and using personal data for crime analysis.
- **Resource Allocation Challenges:** Decisions about resource allocation may not be as data-driven as they could be, resulting in suboptimal use of law enforcement resources.

CHAPTER 2

SYSTEM ARCHITECTURE

2.1. TECHNOLOGY USED

2.1.1. STREAMLIT

Streamlit offers several benefits in the context of the provided crime prediction and reporting web application:

- **Rapid Prototyping and Development:**

Streamlit simplifies the process of creating interactive web applications with minimal code. It allows for rapid prototyping, enabling quick development and testing of different features.

- **Intuitive User Interface:**

Streamlit provides an easy-to-use interface with simple elements like sliders, buttons, and text inputs. This makes it accessible for users without extensive programming experience, facilitating intuitive incident reporting.

- **Interactive Data Visualization:**

Streamlit seamlessly integrates with visualization libraries like Folium, allowing for interactive maps and data visualizations. This enhances the user experience by providing an intuitive way to explore and understand crime data.

- **Real-Time Data Updates:**

The reactive nature of Streamlit enables real-time updates without the need for manual refreshing. In the incident reporting feature, users can see immediate feedback after reporting an incident, enhancing user engagement.

- **Efficient Data Handling:**

Streamlit simplifies the integration with Pandas and Geopandas for efficient data handling and manipulation. This is crucial for processing and presenting crime data in a meaningful way.

- **Seamless Deployment:**

Streamlit applications can be easily deployed and shared through various platforms. This simplifies the process of making the crime prediction and reporting tool accessible to a broader audience.

- **Integration with Machine Learning Models:**

If advanced crime prediction models are developed, Streamlit can seamlessly integrate them into the web application. This allows for a holistic solution that combines incident reporting, visualization, and predictive analytics.

- **Community Support and Documentation:**

Streamlit has a growing and active community, providing support and a wealth of documentation. This can be valuable for developers seeking assistance and resources during the development process.

- **Open Source and Customization:**

Streamlit is an open-source framework, allowing developers to customize and extend its functionality according to project requirements. This flexibility is beneficial for tailoring the web application to specific needs.

- **Integration with External Services:**

In the provided example, Streamlit is integrated with Twilio for incident reporting. This showcases Streamlit's capability to seamlessly connect with external services, expanding the application's functionality.

2.1.2 Algorithm Used

K-Means Clustering in the Crime Prediction and Reporting Web Application:

Overview:

K-Means clustering is an unsupervised machine learning algorithm that partitions a dataset into 'k' distinct, non-overlapping subsets (clusters). Each data point belongs to the cluster with the nearest mean, and the algorithm aims to minimize the variance within each cluster.

Application in the Web App:

In the crime prediction and reporting web application, K-Means clustering is applied to the latitude and longitude coordinates of reported crime incidents. The primary goal is to identify major possible crime occurrence areas based on the spatial distribution of reported incidents.

Implementation:

User Interaction: The user can interactively choose the number of clusters ('k') using a Streamlit slider. This flexibility allows users to experiment with different cluster configurations.

K-Means Algorithm:

The KMeans class from the sklearn.cluster module is employed to perform K-Means clustering on the latitude and longitude columns of the crime dataset.

Color-Coded Clusters on the Map:

Each cluster is represented by a unique color on the map, enhancing the visual distinction between different areas.

Outcome:

The result is a map that visually separates the reported incidents into distinct clusters. Users can observe areas with similar crime patterns, providing valuable insights for both law enforcement and the general public. Adjusting the number of clusters allows for flexibility in identifying different levels of granularity in the spatial distribution of crime incidents.

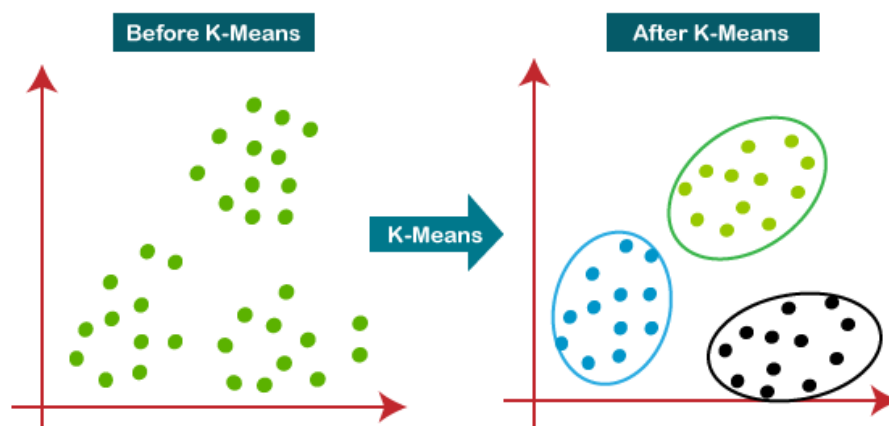


Fig.2.1. K-means diagram

HeatMap in the Crime Prediction and Reporting Web Application:

Overview:

A HeatMap is a graphical representation of data where values are depicted using colors. In the context of the crime prediction and reporting web application, a HeatMap is used to visualize the intensity of reported crime incidents at various geographic points.

Application in the Web App:

The HeatMap is employed to dynamically illustrate crime hotspots on the map. It offers a visual representation of the concentration of reported incidents, with color intensity indicating areas with higher crime frequencies.

Implementation:

Creation of HeatMap:

The folium library is utilized to create an interactive HeatMap on the geographical map. The HeatMapWithTime class is used to represent the temporal evolution of crime incidents.

Dynamic Visualization:

The HeatMap provides a dynamic and interactive way to explore the temporal patterns of reported incidents. Users can adjust the time slider to view how crime hotspots evolve over time.

Adjustable Radius:

The radius parameter in the HeatMapWithTime class is adjustable, allowing users to control the granularity of the HeatMap representation.

Outcome:

The HeatMap provides a visually intuitive representation of crime hotspots over time. Users can observe areas with consistently high crime incident frequencies, helping law enforcement and the public gain insights into temporal patterns and spatial concentrations.

Key features and applications of the Crime Prediction and Hotspot mapping:

Accuracy:

Machine learning algorithms can achieve high levels of accuracy in predicting crime rates and identifying crime hotspots.

Scalability:

Machine learning algorithms can be scaled to handle large and complex crime datasets.

Flexibility:

Machine learning algorithms can be used to predict a wide variety of crime types, including both property crimes and violent crimes.

Explainability:

Some machine learning algorithms are able to explain their predictions, which can help users to understand why the algorithm made a particular prediction.

Applications of machine learning for crime prediction and hotspot mapping:

- **Predicting crime rates:**

Machine learning algorithms can be used to predict crime rates for specific geographic areas and time periods. This information can be used by law enforcement and other agencies to develop and implement targeted crime prevention strategies.

- **Identifying crime hotspots:**

Machine learning algorithms can be used to identify areas where crime is most likely to occur. This information can be used by law enforcement and other agencies to deploy resources more effectively and to develop targeted crime prevention strategies.

- **Analyzing crime patterns and trends:**

Machine learning algorithms can be used to analyze crime patterns and trends over time and space. This information can be used to identify the root causes of crime and to develop more effective crime prevention strategies.

The provided Streamlit web application for crime prediction and reporting primarily involves machine learning, specifically the use of the K-Means clustering algorithm. The K-Means algorithm is a machine learning technique categorized under unsupervised learning, as it doesn't rely on labeled training data.

Machine learning is a broader category that encompasses various techniques and algorithms, and K-Means clustering is one such algorithm used for pattern recognition and grouping of data points.

Deep learning typically refers to the use of neural networks with multiple layers (deep neural networks) for more complex tasks, such as image recognition, natural language processing, and other tasks that involve intricate patterns and representations. In the provided code, there isn't explicit usage of deep learning techniques.

If you were to incorporate deep learning into this project, it might involve the development and training of neural network models to predict crime incidents based on historical data. However, such an extension would likely involve a more complex setup, including the design of a neural network architecture, data preprocessing, and training procedures.

Crimes Dataset

This dataset represents the real-world crimes in Tamilnadu. It includes criminal offenses and crime incidents in the city and county of the city).

This dataset is composed of 15 attributes with 333068 instances. The key attributes provide the offense type and its category such as robbery, public-disorder, and sexual assault.

The dataset also gives the exact occurrence time of the crime along with the district, the neighborhood and the exact geographic location. The following table shows the used key attributes and its content values

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	date	time_of_day	crime_type	location	latitude	longitude	victim_gender	victim_age	perpetrator_gender	perpetrator_age	weapon	injury	weather	temperature	previous_activity
2	12/18/2021	21:29:18	Assault	chennai	13.1206	77.5712	Female	62	Other	18	Gun	Major	Partly Cloudy	40	Fraud
3	9/16/2018	8:36:10	Forgery	chennai	13.068	77.5754	Male	56	Male	18	None	Fatal	Partly Cloudy	10	None
4	6/2/2018	4:27:36	Embezzlement	chennai	13.0788	77.6458	Other	18	Male	20	Gun	Fatal	Partly Cloudy	5	Robbery
5	2/24/2020	10:51:00	Robbery	chennai	12.8899	77.5503	Male	25	Male	25	Gun	Minor	Clear	25	Drug Offense
6	6/2/2018	8:01:24	Forgery	chennai	13.0214	77.4496	Male	53	Other	63	Knife	None	Clear	20	Fraud
7	2/24/2018	21:00:41	Embezzlement	chennai	13.0559	77.4998	Other	36	Female	47	Gun	Fatal	Overcast	1	Robbery
8	4/13/2019	18:54:23	Vandalism	chennai	13.0814	77.5721	Other	61	Female	39	Knife	Major	Overcast	0	Burglary
9	4/16/2021	21:58:36	Forgery	chennai	13.0547	77.5186	Female	25	Female	23	Knife	Major	Clear	25	None
10	1/13/2022	3:38:23	Forgery	chennai	12.9809	77.7966	Female	58	Female	18	Gun	Minor	Clear	0	Embezzlement
11	3/5/2018	9:25:17	Burglary	chennai	13.0961	77.4986	Female	35	Other	49	Knife	None	Clear	37	Drug Offense
12	4/29/2021	6:56:17	Theft	chennai	12.9208	77.4237	Other	30	Female	28	None	Fatal	Overcast	20	Burglary
13	8/2/2022	20:16:00	Fraud	chennai	13.1532	77.6746	Female	55	Male	48	Blunt Object	Minor	Clear	7	Forgery
14	7/5/2021	20:33:33	Robbery	chennai	12.9826	77.7132	Other	62	Female	23	None	Fatal	Rain	10	Embezzlement
15	10/1/2018	18:13:30	Vandalism	chennai	13.1391	77.62	Male	27	Female	31	None	None	Clear	33	Assault
16	9/28/2021	6:36:59	Embezzlement	chennai	12.8746	77.7989	Other	37	Female	42	Gun	Major	Rain	8	Fraud
17	6/22/2020	23:05:43	Theft	chennai	12.9363	77.4805	Male	21	Other	61	Blunt Object	Major	Partly Cloudy	17	Drug Offense
18	8/13/2022	0:37:42	Vandalism	chennai	12.9822	77.5382	Male	27	Other	52	Knife	None	Clear	3	Theft
19	5/11/2019	6:28:02	Assault	chennai	13.0582	77.6439	Other	43	Other	58	Knife	Major	Partly Cloudy	15	None
20	7/10/2022	15:26:00	Embezzlement	chennai	13.1544	77.439	Male	27	Female	22	None	Major	Overcast	27	Vandalism
21	11/30/2021	22:34:24	Burglary	chennai	13.0076	77.5068	Other	62	Female	57	Knife	Major	Snow	40	Fraud
22	1/27/2021	4:25:27	Burglary	chennai	12.8912	77.4086	Male	40	Other	61	Blunt Object	Minor	Snow	30	Burglary
23	12/27/2022	15:19:12	Embezzlement	chennai	13.0014	77.4377	Male	33	Other	27	Gun	Major	Overcast	24	Theft
24	2/22/2019	23:52:51	Fraud	chennai	13.1463	77.4136	Other	25	Male	31	Gun	Major	Snow	8	Fraud
25	5/21/2018	19:11:22	Drug Offense	chennai	13.1326	77.4475	Other	46	Other	47	None	None	Partly Cloudy	11	Vandalism
26	4/14/2019	21:19:32	Assault	chennai	13.1024	77.5917	Other	18	Other	43	Knife	None	Snow	28	Fraud

Fig.2.2 Crime Dataset

Evaluation

In this section, we evaluate each of the three constructed models regarding different aspects. Apriori Algorithm is our first model that was used to extract frequent crime patterns. The key strength for this model are its readiness and easiness of use and implement.

On the contrary, the main drawback is slowness. It takes very long running time to give the results, especially with smaller values of the minimum support. The other two models are Naïve Bayesian and decision tree classifiers that were used for crime type prediction.

We applied the 5-fold cross validation strategy on both models then compared the prediction accuracy for each city. Regarding to the Naïve Bayesian classifier, it achieves an accuracy of 51% in chennai crime prediction while it reaches 54% for Tamilnadu crime prediction. On the other hand, decision tree classifier reports less prediction accuracy with 42 % for Chennai and 43% for Tamilnadu. Moreover, the decision tree model created a very complex tree that cannot generalize the data for both cities. However, the two classifiers have the same performance in terms of their running time. Since the Bayesian classifier yields the best overall performance, we chose it as the ideal model for crime prediction in our study. Table reports the confusion matrix of applying this model on chennai testing set. That shows a report of the main classification metrics obtained by this model.

2.2. IMPLEMENTATION DETAILS

Data collection: The crime data can be collected from a variety of sources, such as police reports, victim surveys, and census data. The data can be collected using a variety of methods, such as web scraping, API calls, and manual data entry.

Data preprocessing: Once the data has been collected, it needs to be preprocessed to clean it and make it suitable for machine learning. This may involve removing outliers, imputing missing values, and converting the data into a format that is compatible with the chosen machine learning algorithm.

Feature engineering: Once the data has been preprocessed, features need to be engineered. This involves transforming the data into features that are informative for the machine learning algorithm. Some common features that are used for crime prediction include:

- Time of day: The time of day when the crime occurred.
- Day of week: The day of the week when the crime occurred.
- Location: The geographic location where the crime occurred.
- Type of crime: The type of crime that occurred.

Environmental factors: Factors such as poverty, unemployment, and lack of education.

Model training: Once the features have been engineered, the machine learning model can be trained. This involves feeding the model the training data and allowing it to learn the relationships between the features and the target variable (crime rate).

Model evaluation: Once the model has been trained, it needs to be evaluated to assess its performance. This can be done by feeding the model the test data and seeing how well it predicts the crime rates for these areas.

Model deployment: Once the model has been evaluated and found to be performant, it can be deployed to production. This may involve integrating the model into a software application or making it available as a web service.

Hotspot mapping: The output of the machine learning model can be used to create hotspot maps. Hotspot maps are visualizations that show the areas where crime is most likely to occur.

To create a hotspot map, the crime predictions from the machine learning model can be aggregated to the neighborhood or block level. The aggregated crime predictions can then be visualized using a MAPS software application.

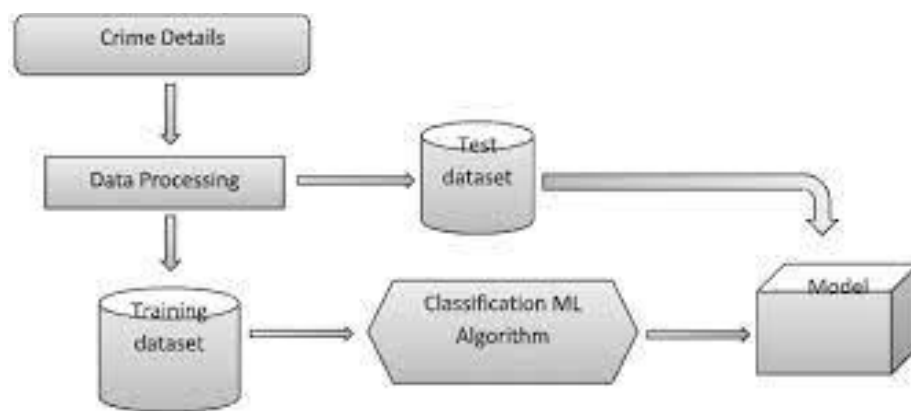


Fig.2.3 Flow Chart

Here are some specific implementation details for each component of the architecture:

Data collection: The crime data can be collected using a variety of methods, such as:

Web scraping: The crime data can be scraped from public websites, such as police department websites and news websites.

API calls: The crime data can be accessed using APIs that are provided by law enforcement agencies and other organizations.

Manual data entry: The crime data can be manually entered into a database.

Data preprocessing: The data preprocessing steps may vary depending on the specific machine learning algorithm that is being used. However, some common preprocessing steps include:

Removing outliers: Outliers are data points that are significantly different from the rest of the data. Outliers can skew the results of the machine learning algorithm, so they should be removed before the model is trained.

Imputing missing values: Missing values can occur in the data for a variety of reasons. Missing values can be imputed using a variety of methods, such as mean imputation, median imputation, and random forest imputation.

Converting the data into a format that is compatible with the chosen machine learning algorithm: The data needs to be converted into a format that is compatible with the chosen machine learning algorithm. For example, some machine learning algorithms require the data to be in a numerical format.

Feature engineering: The feature engineering steps may also vary depending on the specific machine learning algorithm that is being used.

Some common feature engineering techniques include:

Creating new features from existing features: New features can be created from existing features by combining features in different ways or by transforming the features using mathematical operations.

Using statistical methods to reduce the dimensionality of the data: The dimensionality of the data can be reduced using statistical methods such as principal component analysis (PCA). Reducing the dimensionality of the data can help to improve the performance of the machine learning algorithm.

Model training: The machine learning model can be trained using a variety of different machine learning libraries, such as TensorFlow, PyTorch, and scikit-learn. The specific training steps will vary depending on the chosen machine learning algorithm.

Model evaluation: The model evaluation steps will also vary depending on the chosen machine learning algorithm.

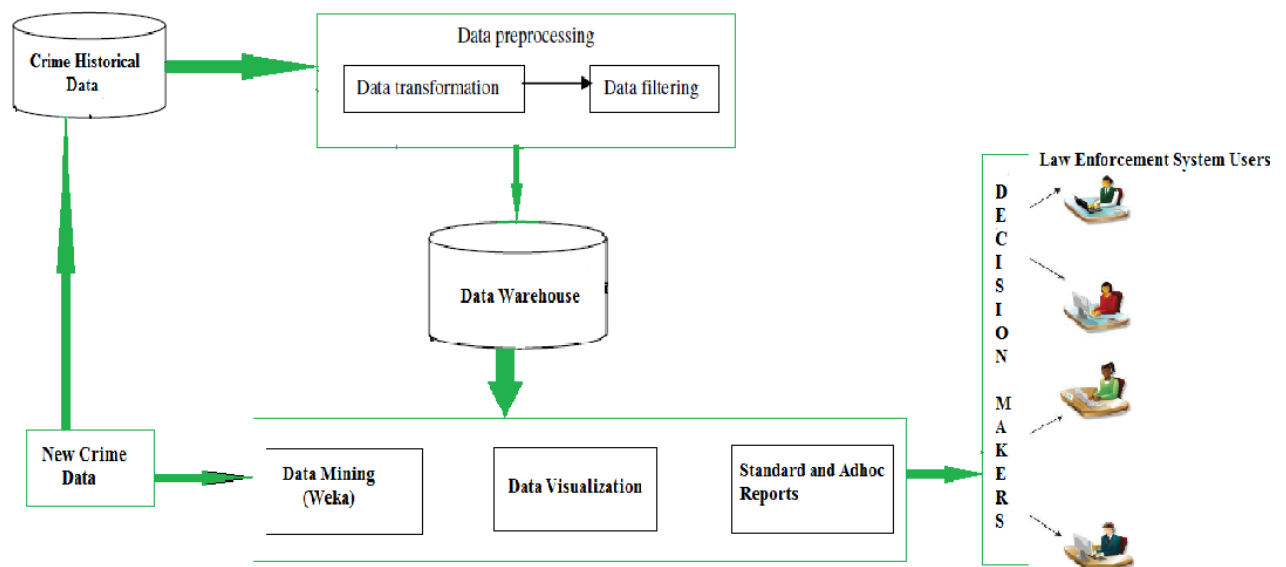


Fig.2.4 Architecture of Data processing

Some common evaluation metrics include:

Accuracy: Accuracy measures the proportion of predictions that the model got correct.

Precision: Precision measures the proportion of positive predictions that were actually correct.

Recall: Recall measures the proportion of actual positives that the model predicted correctly.

2.3 MODULES

Input parameter

- Date
- Time_of_day
- Crime_type
- Location
- Latitude
- Longitude
- Gender

2.4. USER INTERFACE (UI)

Description: The UI Module is the user-facing component of your tool, which simplifies web application development. The UI Module allows users to interact with the tool, input preferences, and visualize crime data.

Functionality: It provides a clean and intuitive interface, where users can interact with data visualizations, input criteria, and customize the appearance of the tool. It acts as a bridge between users and the tool's backend.

Key Considerations: Ensuring a user-friendly interface and responsive design for different devices and screen sizes is essential.

Data Visualization Module

Description: The Data Visualization Module is responsible for creating visual representations of crime data. It integrates with data visualization libraries such as Matplotlib, Plotly, and Altair to generate dynamic charts, graphs, and tables.

Functionality: Users can explore and interpret financial data through interactive visualizations. This module translates structured data into meaningful and visually engaging representations.

Key Considerations: Selecting appropriate visualization techniques and ensuring that visualizations are interactive and informative are critical.

Train Model

The preprocessed data is then used as input for the Deep Learning Model. This Deep Learning model consists of a RNN, STNN and Binary Classifier, which will be used for prediction of Crime Hotspots. In RNN, the input data is the crime rate of each cell for a given time interval and the output is the predicted crime rate of each cell. The relu function has been used as the activation function. Finally, the obtained output and compare it with the expected output. Training is carried out until the loss is considerably low with model weights stored at regular intervals. In STNN, the input data is the crime rate of the neighbouring cells of each cell as well as its historical crime rate is used to train the STNN.

The STNN is trained until the loss is considerably low. After this, the crime rate of each cell will undergo a Binary Classification to predict whether the cell is a 'Hotspot' or not.

Benefits of using Machine Learning

There are many benefits to using machine learning for crime prediction and hotspot mapping. Some of the key benefits include:

- **Improved crime prevention:** By identifying crime hotspots and predicting where crime is most likely to occur, law enforcement agencies can deploy resources more effectively and develop targeted crime prevention strategies.
- **Reduced crime rates:** Studies have shown that machine learning can be effective in reducing crime rates. For example, a study by the University of Chicago found that PredPol, a machine learning-based crime prediction system, was able to reduce crime rates by up to 15% in the areas where it was deployed.
- **Enhanced public safety:** By providing law enforcement agencies with better information about crime, machine learning can help to enhance public safety and make communities safer.
- **Increased transparency and accountability:** Machine learning algorithms can be transparent and accountable, allowing law enforcement agencies to explain how they are making decisions about crime prediction and hotspot mapping.

In addition to these benefits, machine learning can also be used to:

- **Identify trends and patterns in crime data:** Machine learning can be used to identify trends and patterns in crime data that would be difficult or impossible to identify manually. This information can be used to develop more effective crime prevention strategies.
- **Evaluate the effectiveness of crime prevention programs:** Machine learning can be used to evaluate the effectiveness of crime prevention programs by tracking crime rates before and after the programs are implemented.

- Identify areas where social services are needed: By identifying areas where crime is concentrated, machine learning can help to identify areas where social services are needed. This information can be used to develop targeted social programs to address the root causes of crime.

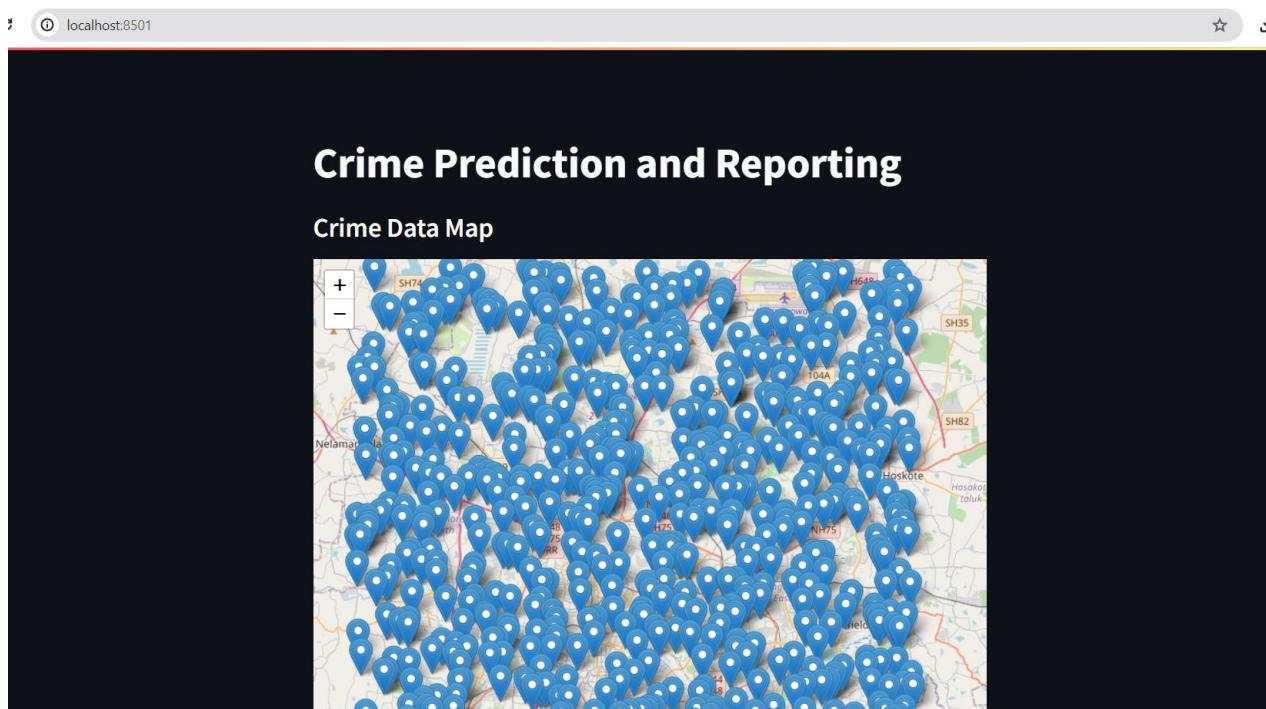


Fig.2.5 Crime predict and mapping

```
report = classification_report(y_test, y_pred)
print(report)
```

	precision	recall	f1-score	support
0	0.16	0.20	0.18	15
1	0.04	0.12	0.06	8
2	0.12	0.06	0.08	17
3	0.14	0.31	0.20	13
4	0.12	0.07	0.09	14
5	0.00	0.00	0.00	14
6	0.17	0.24	0.20	21
7	0.11	0.05	0.07	21
8	0.00	0.00	0.00	17
accuracy			0.11	140
macro avg	0.10	0.12	0.10	140
weighted avg	0.10	0.11	0.10	140

Fig.2.6 Accuracy and average

Report an Incident

Your Location (latitude, longitude)

2023/11/17

Location

Your Location (latitude, longitude)

Incident Type

Assault

Time

Victim_Gender

Report

Fig.2.7 Report an Incident

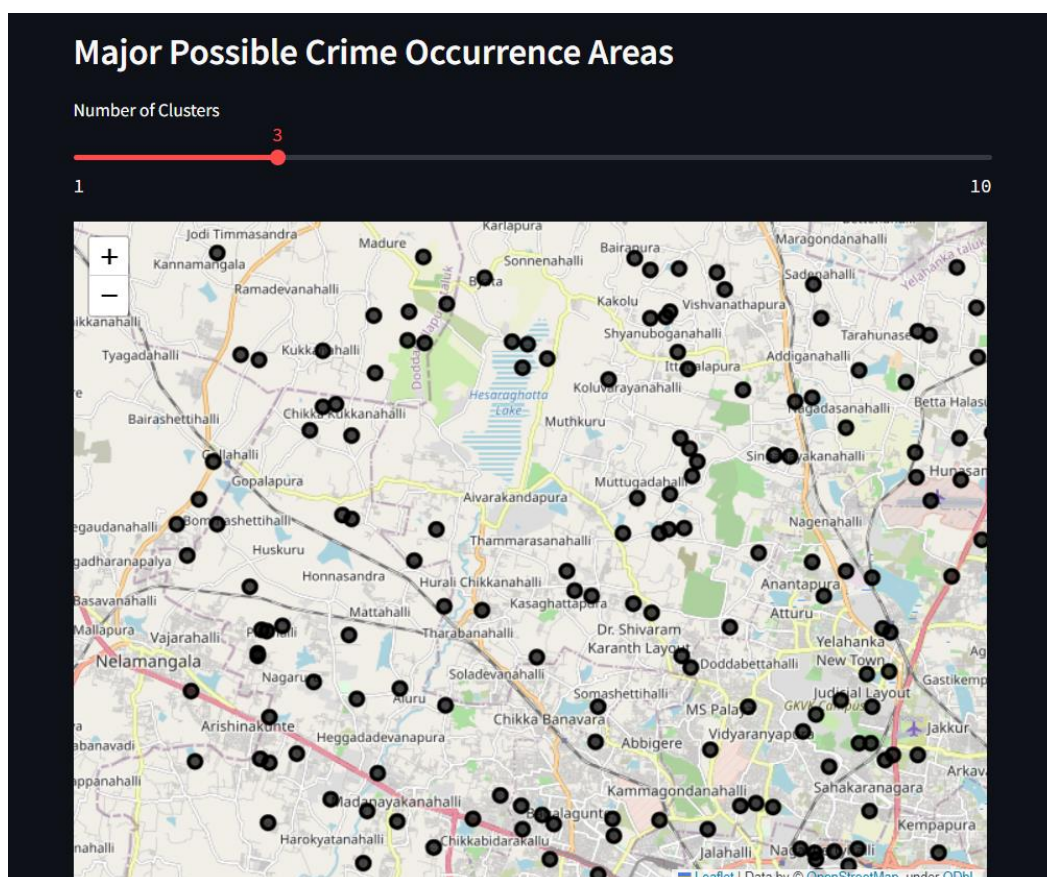


Fig.2.8 Major Possible Crime

Location Data Table

	location	latitude	longitude	crime_type	↓ cluster
728	saalem	47	45	Assault	2
723	saalem	27	52	Forgery	2
722	None	42.5195	72.5	Assault	2
721	asdw	42.5195	72	Assault	2
705	jammu and kash	33.7782	76.5762	Drug Offense	2
700	saalem	43	106	Assault	2
727	sona college	11	22	Burglary	1
726	sona college	11	22	Burglary	1
725	saalem	15	79	Vandalism	0
724	PODILI	15.607	79.6146	Robbery	0

Fig.2.9 Location Data Table

CHAPTER 3

CONCLUSION AND FUTURE WORKS

CONCLUSIONS

In the realm of crime prediction and hotspot mapping, the integration of data-driven systems and advanced technologies has proven to be a powerful tool for law enforcement agencies. The development of predictive models, such as the Random Forest Classifier, along with the incorporation of Natural Language Processing (NLP) and other preprocessing techniques, enables agencies to analyze historical crime data, identify patterns, and make proactive decisions to enhance public safety.

The utilization of geographic information systems (GIS) and spatial analysis techniques has allowed for the creation of detailed hotspot maps, enabling law enforcement to focus resources on high-risk areas. By combining historical crime data, spatial factors, demographic information, and environmental data, these systems provide actionable insights that improve resource allocation and crime prevention strategies.

However, there are challenges and considerations that must be addressed. Ethical and privacy concerns, data quality, and the need for continuous model refinement remain important aspects of system development. Additionally, effective communication with the community and adherence to legal standards are crucial to ensure the trust and cooperation of the public.

Future Works:

As technology and data science continue to advance, the field of crime prediction and hotspot mapping holds promise for several future developments:

- **Real-Time Analysis:** Enhancing systems to provide real-time crime prediction and hotspot mapping will enable law enforcement agencies to respond more effectively to dynamic crime patterns.

- **Machine Learning Advancements:** Leveraging advanced machine learning techniques, such as deep learning and neural networks, may provide even more accurate predictions by capturing complex patterns in crime data.
- **Integration of IoT:** Incorporating data from Internet of Things (IoT) sensors and surveillance cameras can offer valuable insights for crime prediction and investigation.

REFERENCES

1. Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, 21(1-2), 4-28.
2. Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2011). Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, 106(493), 1110-1124.
3. Chainey, S., & Ratcliffe, J. H. (2005). *GIS and crime mapping*. John Wiley & Sons.
4. Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493), 100-108.
5. Bernasco, W., & Block, R. (2011). Robberies in Chicago: a block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. *Journal of Research in Crime and Delinquency*, 48(1), 33-57.
6. Tita, G. E., Ridgeway, G., & Abrahamse, A. (2006). Criminal mobility and crime prevention. *Journal of Quantitative Criminology*, 22(2), 129-153.
7. Weisburd, D., & Mazerolle, L. G. (2000). Crime and disorder in drug hot spots: Implications for theory and practice in policing. *Police Quarterly*, 3(3), 331-349.
8. Andresen, M. A., Malleson, N., & Tompson, L. (2018). Crime seasonality and its variations across space. *Applied Geography*, 91, 31-39.
9. Mohler, G. O., Short, M. B., Brantingham, P. J., & Schoenberg, F. P. (2012). Bayesian space-time integration in the modeling of random processes. *Journal of the American Statistical Association*, 107(500), 196-207.
10. Townsley, M., Homel, R., & Chaseling, J. (2003). Infectious burglaries. A test of the near repeat hypothesis. *British Journal of Criminology*, 43(3), 615-633.
11. <https://www.fbi.gov/how-we-can-help-you/more-fbi-services-and-information/ucr>

12. <https://www.esri.com/en-us/industries/law-enforcement/>
13. <https://www.crimemapping.com/>
14. <https://www.policinginstitute.org/>

APPENDIX -I

Main.py

```
import streamlit as st
import pandas as pd
import geopandas as gpd
import folium
from streamlit_folium import folium_static
from sklearn.cluster import KMeans
from twilio.rest import Client

crime_data=pd.read_csv('crime.csv')

gdf=gpd.GeoDataFrame(crime_data,
geometry=gpd.points_from_xy(crime_data.longitude, crime_data.latitude))

TWILIO_ACCOUNT_SID="
TWILIO_AUTH_TOKEN="
TWILIO_PHONE_NUMBER="
POLICE_PHONE_NUMBER="

st.title("Crime Prediction and Reporting")

st.subheader("Crime Data Map")
m=folium.Map(location=[crime_data['latitude'].mean(),
crime_data['longitude'].mean()], zoom_start=12)
for idx,row in gdf.iterrows():
    folium.Marker([row['latitude'], row['longitude']],
popup=row['crime_type']).add_to(m)
folium_static(m)

st.subheader("Report an Incident")
date=st.date_input("Your Location (latitude, longitude)")
location=st.text_input("Location")
user_location=st.text_input("Your Location (latitude, longitude)")
incident_type=st.selectbox("Incident Type", crime_data['crime_type'].unique())
time=st.text_input("Time")
gender=st.text_input("Victim_Gender")
report_button=st.button("Report")
```

```

if report_button:
    if user_location:
        user_location = user_location.split(',')
        if len(user_location) == 2:
            user_latitude, user_longitude = float(user_location[0]),
            float(user_location[1])
            new_report = {
                'date': date,
                'latitude': user_latitude,
                'longitude': user_longitude,
                'crime_type': incident_type,
                'time_of_day': time,
                'location': location,
                'victim_gender': gender,
            }
            try:
                twilio_client = Client(TWILIO_ACCOUNT_SID,
                TWILIO_AUTH_TOKEN)
                message = twilio_client.messages.create(
                    body=f"Emergency: Crime reported at {location}. Type:
                    {incident_type} date: {date} latitude: {user_latitude} longitude: {user_longitude}
                    .",from_="+13344630998,to="+919361117846)
                st.success("Incident reported, and SMS to police sent successfully!")
            except Exception as e:
                st.error(f"Error sending SMS to police: {str(e)}")

            crime_data = pd.concat([crime_data, pd.DataFrame([new_report])],
            ignore_index=True)
            crime_data.to_csv('crime.csv', index=False)

            st.success("Incident reported and saved to the CSV file.")

st.subheader("Major Possible Crime Occurrence Areas")
k=st.slider("Number of Clusters", 1, 10, 3)
kmeans = KMeans(n_clusters=k, random_state=0)
crime_data['cluster'] = kmeans.fit_predict(crime_data[['latitude', 'longitude']])
clustered_map=folium.Map(location=[crime_data['latitude'].mean(),
crime_data['longitude'].mean()], zoom_start=12)

for cluster in crime_data['cluster'].unique():
    cluster_data=crime_data[crime_data['cluster'] == cluster]
    color = "#{:02x}{:02x}{:02x}".format(int(255 * (cluster / k)), 0, 0)
    for idx, row in cluster_data.iterrows():
        folium.CircleMarker([row['latitude'], row['longitude']],

```

```
radius=5,  
color=color,  
fill=True,  
fill_color=color,  
fill_opacity=0.7,  
popup=f"Cluster {cluster}").add_to(clustered_map)
```

```
folium_static(clustered_map)
```

```
st.subheader("Location Data Table")
```

```
st.write(crime_data[['location', 'latitude', 'longitude', 'crime_type', 'cluster']])
```