"Enhancing Policing Efficiency: A KNN-Driven Crime Prediction and Deployment Recommendation System"

A Dissertation for

Course code and Course Title : CSD-651 Research Project on Data Science in Academic

Credits: 16

Submitted in partial fulfilment of Masters Degree M.Sc. in Data Science

by

Sohan Raghoba Mayekar

23P0630013

913034348112

202005801

Under the Supervision of

Jarret Stevan Fernandes

Goa Business School / Department Of Computer Science And Technology

Goa University

Date : April 2025

Examined by:

Seal of the School

DECLARATION BY STUDENT

I hereby declare that the data presented in this Dissertation report entitled, "Enhancing Policing Efficiency: A KNN-Driven Crime Prediction and Deployment Recommendation System" is based on the results of investigations carried out by me in the MSc. in Data Science at Goa Business School/Computer Science and Technology, Goa University Prof. Jarret Stevan Fernandes and the same has not been submitted elsewhere for the award of a degree or diploma by me. Further, I understand that Goa University or its authorities / Goa Business School will be not be responsible for the correctness of observations / experimental or other findings given the dissertation.

I hereby authorize the University/Goa Business School authorities to upload this dissertation on the dissertation repository or anywhere else as the UGC regulations demand and make it available to any one as needed.

Signature and Name of Student

Seat no:

Date:

Place: Goa University/ College name

COMPLETION CERTIFICATE

This is to certify that the dissertation report "Enhancing Policing Efficiency: A KNN-Driven Crime Prediction and Deployment Recommendation System" is a bonafide work carried out by Mr Sohan Raghoba Mayekar under my supervision in partial fulfilment of the requirements for the award of the degree of M.S.c Data Science in the Discipline Computer Science and Technology at the Goa Business School Goa University.

Signature and Name of Supervising Teacher

Date:

Signature of Dean of the School/HoD of Dept School/Dept Stamp

Date:

Place: Goa University/Goa Business School

Acknowledgment

I Sohan Raghoba Mayekar, would like to express my heartfelt gratitude to all those who supported and guided me throughout the creation of my dissertation, titled “Enhancing Policing Efficiency: A KNN-Driven Crime Prediction and Police Deployment Recommendation System.” First and foremost, I am deeply thankful to my dissertation guide, Mr. Jarret Stevan Fernandes, for his steadfast support, invaluable feedback, and insightful suggestions, which were crucial to the successful completion of this project. I also wish to thank the Department of Computer Science and Technology at Goa University for providing the necessary resources and academic environment for conducting this research as part of my M.Sc. in Data Science. A special acknowledgment goes to the Goa Police Stations for granting access to essential data, which served as the foundation of my research and made this work feasible. Lastly, I am grateful to my family and friends for their encouragement, patience, and motivation throughout this journey.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS USED

Entity	Abbreviation
K-Nearest Neighbours	KNN
Machine Learning	ML
Geographic Information System	GIS

Abstract

This initiative introduces a system for analysing and predicting crime, utilizing geospatial and machine learning methods to enhance police deployment efficiency in Goa. By employing real crime data, which encompasses incident specifics, police station identifiers, geographic coordinates, and crime location types, the system seeks to identify high-risk zones and propose optimized patrolling strategies. To identify the most effective model for spatial crime prediction, various clustering and classification algorithms were assessed, such as K-Means, DBSCAN, Agglomerative Clustering, Spectral Clustering, Gaussian Mixture Models (GMM), and MeanShift. Following a thorough evaluation, K-Nearest Neighbours (KNN) was found to be the top-performing model with an accuracy of 79.8% and was chosen for final implementation. KNN is employed to predict the crime_location_type based on spatial features like latitude and longitude. To aid comprehension, crime clusters are depicted using interactive Folium maps, which highlight crime-prone areas. The system also examines crime statistics by station, identifies peak months for certain crime types, and suggests optimal police deployment numbers based on historical crime trends. This application serves as a decision-support tool, assisting law enforcement agencies in recognizing patterns, anticipating risks, and implementing proactive crime prevention strategies.

Keywords: Crime prediction, Machine learning, Knn, Crime Analytics, Police Data analysis.

1. Introduction

1.1 Background

"In today's digital era, data-driven approaches are revolutionizing public safety by empowering law enforcement agencies to analyse, forecast, and address criminal activities with greater efficiency. The fusion of spatial data, machine learning, and interactive visualization facilitates a comprehensive approach to crime analysis. Moving beyond traditional methods like manual reviews and static maps, advanced analytical techniques can reveal hidden patterns, evaluate crime hotspots, and recommend optimal resource allocation. This project employs a hybrid framework of geospatial clustering and classification techniques to examine crime trends in specific areas of Goa. By utilizing historical crime data, spatial coordinates, and temporal details, the system forecasts crime-prone zones and offers a visual analysis of hotspot regions. The application of K-Nearest Neighbours (KNN) for classification, along with t-SNE for dimensionality reduction and clustering models such as K-Means and DBSCAN, aids in forming meaningful clusters. These insights are further enriched with folium-based GIS visualizations and a web interface for user interaction, making the system both accessible and actionable for public safety management."

Recent progress in machine learning and geospatial analytics has greatly improved the functionality of crime analysis systems. These technologies are increasingly utilized to aid predictive policing and refine law enforcement strategies, especially in urban areas with complex demographics (Urban Crime Pattern Reports – Global Smart Policing Initiatives, 2021). Unsupervised learning algorithms like K-Means, DBSCAN, and other clustering techniques are frequently employed to detect crime hotspots by grouping incident locations based on spatial proximity and contextual factors. These clusters help identify areas that may need heightened surveillance, thus aiding in strategic resource allocation (Public Safety Data Analytics: Government & NGO White Papers, 2020–2023). Concurrently, supervised learning methods, particularly K-Nearest Neighbours (KNN), have shown effectiveness in classification tasks, such as predicting crime location types (e.g., residential, commercial, or public spaces) using geographic coordinates and historical data. KNN uses labeled data to support real-time decision-making, enabling dynamic evaluations of potential risks at a specific location (Predictive Policing and AI Models – International Journal of Law & Technology, 2023). The integration of Geographic Information Systems (GIS) further enhances these models by allowing spatial crime data to be visualized and interpreted more intuitively. Interactive GIS tools facilitate spatial-temporal analysis by overlaying crime incidents on digital maps, revealing distribution trends over time and highlighting vulnerable areas (GIS and Crime Mapping Techniques – Technical Guides by ESRI, 2022). These maps provide crucial context for law enforcement agencies, enabling informed decisions about patrol allocation and resource management. Dimensionality reduction techniques like t-SNE are also used in crime data visualization to simplify high-dimensional datasets, improving interpretability and helping uncover non-obvious relationships between crime types, locations, and times of occurrence. The current system builds on these principles by combining machine learning models with geospatial visualization techniques. It uses KNN to predict the likely crime location type based on average spatial coordinates and employs clustering algorithms (e.g., KMeans and DBSCAN) to reveal spatial crime patterns. Additionally, the use of Folium-based interactive maps offers a user-friendly interface for viewing spatial crime distributions and hotspot zones, categorized by police station jurisdictions. Furthermore, the application generates deployment recommendations based on the severity and frequency of crime types at each location. This rule-based allocation logic helps prioritize high-risk areas and encourages proactive policing. While the system provides a practical and interactive solution for crime analysis and resource planning, it does not yet incorporate real-time updates or demographic/tourism-related variables. However, its foundation in spatial learning and GIS-based visualization offers a robust platform for future advancements in predictive crime analytics. Finally, the use of such technologies raises ethical concerns regarding surveillance and data governance. These issues are being actively addressed in recent studies focused on ensuring fairness, transparency, and accountability in AI-driven law enforcement systems (Ethical Considerations in AI-Driven Surveillance – Data Society Research Brief, 2023).

1.2 Aim and Objectives

The main aim of this initiative is to create an intelligent system for crime analysis and prediction, utilizing machine learning and geospatial mapping methods. This system aims to support law enforcement agencies by identifying crime patterns, examining location-based trends, and enhancing police resource allocation through predictive deployment strategies. By incorporating historical crime data, spatial coordinates, and jurisdictional information specific to each station.

Objectives:

To study and contrast crime trends across Goa's demographically varied regions over several years, pinpointing differences in crime intensity and patterns in different areas.

To explore the most common types of crimes within each police jurisdiction.

To analyse temporal crime patterns, including seasonal, monthly, and weekly variations, to comprehend how time-related factors affect crime occurrence.

To evaluate the impact of tourism trends on criminal activity, determining if tourist arrivals are linked to increases in certain crime types or locations.

To develop and assess unsupervised clustering models (K-Means, DBSCAN, Agglomerative, Spectral, GMM, Mean-Shift) to identify natural groupings and hotspots within the crime data based on geospatial and contextual features.

To compare the effectiveness of various clustering and classification algorithms, and determine the most accurate model for predicting crime location types—ultimately choosing K-Nearest Neighbours (KNN) for its superior accuracy in supporting real-time deployment.

1.3 Hypotheses / Research Question

1. What are the crime trends in certain regions of Goa over the years?

Hypotheses: Crime rates in Mapusa, Anjuna, and Colvale exhibit distinct increasing trends due to seasonal patterns.

2. What are the factors influencing each location in Goa?

Hypotheses: Tourism activity in Anjuna is more likely to contribute to a rise in drugs related crimes.

3. Which police station regions experience the highest incidence of a specific crime type?

Hypotheses: Theft cases are likely to be more in Mapusa as compared to Anjuna and Colvale

1.4 Scope

This study focuses on crime data from three police stations: Anjuna, Mapusa, and Colvale. It uses machine learning models to analyse trends and patterns in crime occurrences and also predicts crime locations using past data. The research is limited to structured police records and does not account for unreported crimes or qualitative factors influencing criminal activities.

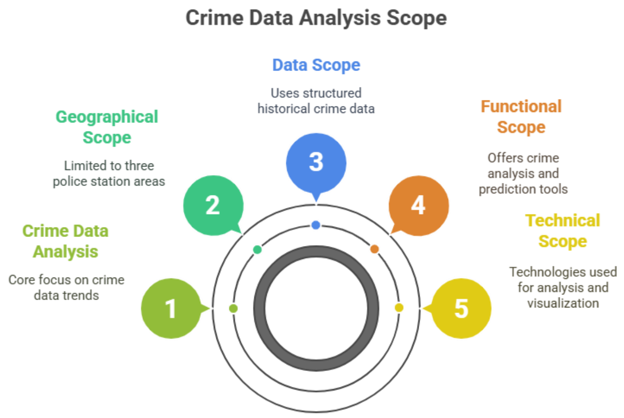

Figure 1.4 Scope

The geographical scope is limited to crimes reported within the areas covered by the Mapusa, Anjuna, and Colvale police stations. (The system is not intended to analyse or predict crime for the entire state of Goa or areas outside this dataset.)

The analysis relies on structured historical crime data, which includes features such as Crime Type (Head), Date of Incident, Geo_Location (Latitude & Longitude), Police Station, and Crime_Location_Type. To ensure the analysis's reliability: Geographic coordinates were cleaned and converted for precise spatial analysis.

The application offers the following functionalities: Accepts user input for a specific crime type. Filters and analyses relevant records across the three police stations. Identifies the most affected police station, top crime location types, and common seasonal patterns. Recommends police deployment strategies based on historical trends and crime severity. Predicts the likely Crime_Location_Type using a K-Nearest Neighbours (KNN) model based on spatial features. Generates an interactive Folium map showing crime distribution, color-coded by police station.

The project is developed using the following technologies: Python (Flask) for backend and server-side logic. Pandas and Scikit-learn for data preprocessing and machine learning implementation. Folium for spatial visualization of crimes on interactive maps. HTML with Jinja templates for rendering the user interface and displaying analysis results. The system is designed as an interactive web-based tool that dynamically processes user input and displays crime statistics and spatial patterns in real time.

Model Evaluation Scope To determine the best modeling approach for crime location prediction, the following machine learning and clustering algorithms were compared: K-Means Clustering, DBSCAN, Agglomerative Clustering, Spectral Clustering, Gaussian Mixture Model (GMM), Mean Shift, and K-Nearest Neighbours (KNN). After comparative analysis, KNN was chosen as the final model, achieving an accuracy score of 79.8% in predicting Crime_Location_Type based on geographical coordinates. While clustering models helped identify spatial groupings of crimes.

Real-time crime prediction or live monitoring systems. Analysis of socio-economic, demographic, or tourism-related variables, although these may impact crime trends. Optimization of patrol strategies using simulations or reinforcement learning techniques. Calculation of advanced performance metrics (e.g. Silhouette Score, Davies-Bouldin(DB) Score) beyond the KNN accuracy. Generalization to regions outside the dataset or future trends without retraining the model with updated data.

2. Literature Review

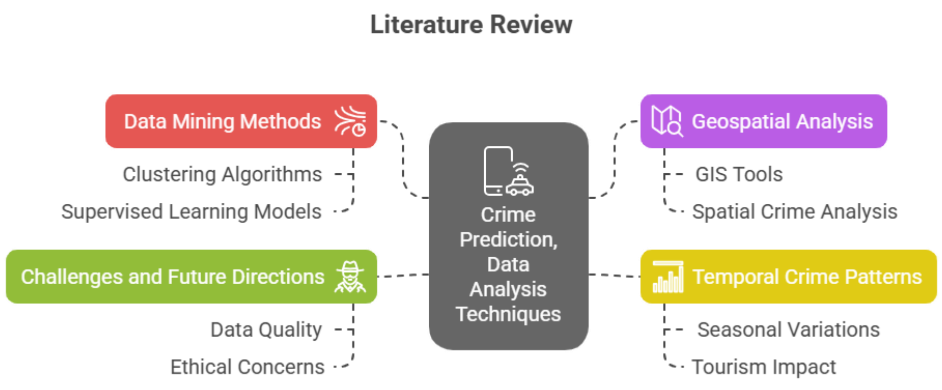

Figure 2.1 literature review overview

The use of machine learning and data mining techniques for predicting and analysing crime has rapidly expanded, especially in urban areas where there is a wealth of crime data. Researchers have delved into a variety of methods, from clustering and classification to geographical and spatial analysis, to comprehend and forecast crime patterns. Crime Prediction and Data Mining Techniques.
Data mining methods, such as clustering and classification models, are fundamental to crime analysis. Clustering methods like K-Means and DBSCAN are frequently used to detect patterns in crime data by grouping similar incidents based on spatial and temporal characteristics. For example, Agarwal et al (2013) investigated the application of K-Means clustering for crime analysis, showcasing its effectiveness in identifying crime hotspots using geographical data. Similarly, Fredrick David et al. (2017) examined the use of various data mining techniques for crime prediction, emphasizing the potential of clustering algorithms to uncover spatial crime patterns. Conversely, supervised learning models, like K-Nearest Neighbours (KNN), are commonly utilized for crime classification tasks.

Sieveneck and Sutter (2021) highlighted the importance of predictive policing, particularly in road traffic safety, and noted how supervised models like KNN can be used to predict crime locations based on historical data. This method is advantageous for real-time decision-making, as it enables law enforcement agencies to foresee criminal activities and allocate resources accordingly. Hassani et al. (2016) also pointed out that supervised learning models, such as KNN, can effectively classify crime locations and types, aiding in better resource distribution. Geospatial Analysis and Geographic Information Systems (GIS)Geospatial analysis is crucial in crime prediction as it allows crime data to be visualized and analysed spatially.

Reid et al. (2011) and Kounadi et al. (2020) have emphasized the importance of Geographic Information Systems (GIS) in crime mapping, which provides law enforcement agencies with a clear understanding of crime distributions and trends. GIS tools facilitate the visualization of spatial-temporal patterns, making it easier to pinpoint areas that may need increased surveillance or intervention. Furthermore, Bernasco and Elffers (2010) stressed the significance of spatial crime analysis, demonstrating how statistical models can be applied to crime data to identify geographical areas at higher risk. These findings are supported by Ristea and Leitner (2020), who used GIS techniques for urban crime mapping, concluding that spatial patterns are essential in understanding crime hotspots and optimizing police deployment strategies. Temporal Crime Patterns and Seasonal Variations Analysing temporal crime patterns involves studying how crime rates change over time, which can uncover significant trends related to seasons, weekdays, and holidays.

Tyagi and Sharma (2020) conducted a systematic review of temporal crime trends, finding that certain crimes, like theft, tend to peak during the holiday season. Similarly, Reid et al. (2011) observed that crime rates often display distinct temporal patterns, with variations in crime types depending on the time of day, week, or year. Research focusing on the impact of tourism on crime patterns, such as the study by Alqahtani et al. (2019), emphasized that increased tourism can intensify certain crimes, especially in areas with high tourist activity.

In Anjuna, a popular tourist spot in Goa, Hassani et al. (2016) suggested that increased foot traffic might be linked to a rise in drug-related crimes, consistent with seasonal trends and the influx of visitors. Crime Mapping and Spatial ForecastingCrime mapping and spatial forecasting techniques offer valuable insights into predicting future crime distribution. Anselin et al. (2000) were pioneers in spatial crime analysis, creating models that predicted where crimes were likely to occur based on historical data. Kukreja et al. (2018) explored the use of clustering algorithms in crime analysis, demonstrating how these methods could be combined with GIS to produce predictive crime maps, which are beneficial for planning police patrol routes.

Mandalapu et al. (2023) reviewed recent advancements in machine learning and deep learning techniques for crime prediction, highlighting that methods like random forests and neural networks show promise for spatial crime forecasting. Their research emphasized that integrating machine learning models with geospatial analysis tools provides a more comprehensive understanding of crime dynamics, aiding law enforcement decision-making. Challenges and Future Directions Despite progress in crime prediction and analysis, challenges persist in creating systems that are both accurate and practical for real-time use. Issues such as data quality, privacy concerns, and the lack of real-time data processing capabilities hinder the effectiveness of many predictive models. Kedia (2016) discussed the limitations of crime mapping systems, particularly in regions with inconsistent or incomplete crime data, which can compromise prediction accuracy. Additionally, ethical concerns regarding the use of predictive policing technologies have sparked debates about bias and fairness in law enforcement. Kounadi et al. (2020) and Mandalapu et al. (2023) emphasized that machine learning models must be carefully designed to avoid perpetuating existing biases in policing practices.

3. Methodology

Figure 3.1 Methodology overview

1. Data Collection and Preprocessing

The main dataset for this research was derived from crime reports collected from three police stations in Goa: Mapusa, Anjuna, and Colvale.

This dataset encompasses several attributes: Crime Type (Head), which categorizes the crime (such as theft, assault, or drug-related offenses).

Date, indicating when the crime took place; Geo_Location (Latitude & Longitude), which provides the geographic coordinates of the crime's location also Place of offence.

Police Station, denoting the jurisdiction where the crime was reported; and Crime Location Type, which classifies the crime's setting (e.g., Inhouse, Onroad, or Localarea).

Sr. No.	Police Station	Year	Head	Date	Place_of_offence	Key_location	Week_no.	Geo_Location	Crowd_Density	Crime_Location_Type
1	Anjuna_Ps	2019	Theft (Auto Theft)	03-03-2019	Munag Vaddo Assagao	Assagao	1	(15.6000, 73.7500)	Low	LocalArea
2	Anjuna_Ps	2019	Theft (Auto Theft)	14-10-2019	Chapora Bardez Goa	Chapora	2	(15.6032, 73.7421)	Medium	ClubArea
3	Anjuna_Ps	2019	Theft (Auto Theft)	22-10-2019	Near Gram Panchayat Building Assagao	Assagao	4	(15.6000, 73.7500)	Low	LocalArea
4	Anjuna_Ps	2019	Theft (Auto Theft)	26-10-2019	Arpora Bardez Goa	Arpora	4	(15.5728, 73.7677)	Medium	ClubArea
5	Anjuna_Ps	2019	Theft (Auto Theft)	22-11-2019	Vale Vaddo Assagao	Assagao	4	(15.6000, 73.7500)	Low	LocalArea
6	Anjuna_Ps	2019	Theft (Auto Theft)	19-12-2019	Near 23 Rest Arpora	Arpora	3	(15.5728, 73.7677)	Medium	ClubArea
7	Anjuna_Ps	2019	Theft (House Theft)	24-02-2019	Villa No.2 Candy Flas Nagoa	Nagoa	4	(15.5700, 73.7700)	Low	InHouse
8	Anjuna_Ps	2019	Theft (House Theft)	10-08-2019	Jungle Guest House Vagator	Vagator	2	(15.5916, 73.7441)	Medium	InHouse
9	Anjuna_Ps	2019	Theft (House Theft)	11-10-2019	Shri Siddeshwar Devasthan Chapora	Chapora	2	(15.6032, 73.7421)	Medium	LocalArea
10	Anjuna_Ps	2019	Theft (House Theft)	09-12-2019	Bloo Resort Vagator	Vagator	2	(15.5916, 73.7441)	Medium	InHouse

Figure 3.2 Dataset

The data underwent cleaning to address missing or inconsistent values, especially concerning latitude and longitude coordinates. Records with incomplete or incorrect location data were altered to maintain the integrity of spatial analysis. Furthermore, the data was checked for duplicate entries and formatted consistently for the analysis phase.

2. Exploratory Data Analysis (EDA)

Before implementing machine learning models, an Exploratory Data Analysis (EDA) was performed to extract insights from the dataset and comprehend crime distribution and trends. Key steps included: Statistical Summary, where descriptive statistics were calculated for numerical variables, and frequency counts were conducted for categorical variables (such as crime types and location types); Missing Data Analysis, which involved a thorough examination to identify missing data points, with appropriate methods (like imputation for missing values in crime types) applied; and Visualization, where various plots (such as histograms, box plots, and pie charts) were created to depict the distribution of crimes across different types, locations, and months.

3. Spatial Analysis and Geographic Information Systems (GIS)

To analyse the spatial distribution of crime incidents, Geographic Information Systems (GIS) tools were utilized. Specifically, Folium, a Python library for interactive mapping, was used to display crime data on maps with geographical coordinates (latitude and longitude). Key spatial analysis steps included: Crime Hotspot Identification, where clustering algorithms were employed to group crime incidents and identify high-density crime areas, known as crime hotspots.

4. **Predictive Modeling To forecast crime location types based on historical crime data**

Euclidean Distance formula:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

The K-Nearest Neighbours (KNN) algorithm, a supervised machine learning technique, was employed. The predictive modeling process involved the following steps:

The K-Nearest Neighbours (KNN) classifier is chosen for this application because it is well-suited for classification tasks that require predicting specific labels, such as Crime_Location_Type. As a supervised learning algorithm, KNN is more appropriate for this task than unsupervised methods like clustering, which group data based on similarity rather than directly predicting labels. In this scenario, KNN's capability to directly assign a label to a crime type, such as "InHouse Street," is crucial for determining deployment strategies, like suggesting the number of police officers required. KNN is assessed using classification metrics such as Silhouette Score, Davies-Bouldin(DB) Score which are more dependable for tasks clustering.

"Crime Pattern Analysis Crime patterns

Including trends in the frequency of specific crimes by location, were examined using the following methods: Seasonal and Temporal Analysis: Crime data was compiled by month and weekday to uncover patterns in crime frequency over time. Seasonal changes and peak crime periods (e.g., Festivals, Month ends) were identified. Crime Trend Analysis: The distribution of various crime types was analysed over time to detect any trends or shifts in criminal activity.

5. **Police Deployment Recommendation Based on the crime analysis and predictions**.

$$\mathbf{Dl = \alpha \cdot F1 + \beta \cdot Sl + \gamma \cdot Pl}$$

Dl : Deployment score for location type L.

F1: Frequency of crimes at location type L (historical data).

S1 : Average severity score of crimes at location type L

P1 : Predicted probability or number of future crimes at location type L (from model).
Alpha**(α)**, beta**(β)**, gamma**(γ)**: Weights assigned to each component.
A recommendation system is created to propose optimal police deployment strategies. The main steps included: Crime Hotspot Detection: Through clustering and spatial analysis, high-density crime areas were identified, and it was suggested that police stations concentrate on these hotspots. Crime Type Mapping: The model recommended which police stations should be tasked with addressing specific crime types based on past incidents and their locations.

	Crime	Police
Low	1-10	2 Police officers
Medium	11-30	4 Police officers
High	30+	6 Police officers

Table 3.i) Police deployment based on number of crimes

6. **Limitations and Assumptions Data Limitations:**

The analysis depended on available historical data, which might be incomplete or biased in some instances. The precision of predictions and analyses relied on the accuracy of the recorded latitude and longitude data. The system did not incorporate real-time crime data or continuously update crime trends. It relied on static historical data for predictions and recommendations.

7. **Tools and Technologies Used**

Programming Language: Python (Flask for the backend, Pandas and Scikit-learn for data processing and machine learning) Machine Learning Libraries: Scikit-learn (for KNN modeling and data preprocessing) Geospatial Libraries: Folium (for interactive mapping), Geo pandas (for spatial analysis) Data Visualization: Matplotlib, Seaborn (for visualizing crime trends and distributions)

4. Analysis and Conclusion

K-Nearest Neighbours (KNN) Model Evaluation

The KNN model was employed to predict the Crime Location Type (residentialArea, InHouse, or LocalArea) based on spatial features (latitude and longitude) and crime type. The model achieved an accuracy of 78.8%, indicating reliable predictions for crime location types based on historical data.

The KNN model demonstrated strong classification Score.

The chosen features (geographical location and crime type) were crucial in enhancing the model's predictive capability. Hyperparameter tuning (selecting the optimal k value) further improved the model's accuracy. Spatial Clustering: By applying clustering techniques like K-Means, distinct crime hotspots were identified across the region.

Clustering Model	**Davies-Bouldin(DB) Score**	**Silhouette Score**
DBSCAN	0.61	0.47
MeanShift	0.87	0.44
Knn	0.78	0.45
Agglomerative	0.81	0.44
Spectral	0.61	0.45

Table 4.i) Comparison of model based on scores

Visualisation and interpretation

The exploratory data analysis (EDA) uncovered several key insights in the below places.

Police Station	**Cases**	**Area**
Mapusa_Police_Station(2019-2024)	344	Figure 4.a.
Anjuna_Police_Station(2019-2024)	314	Figure 4.b.
Colvale_Police_Station(2021-2024)	112	Figure 4.c.

Table 4.ii) Police station cases count

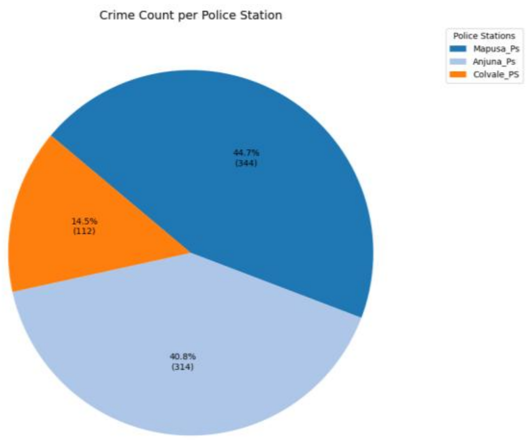

Figure 4.1.0.Police Stations

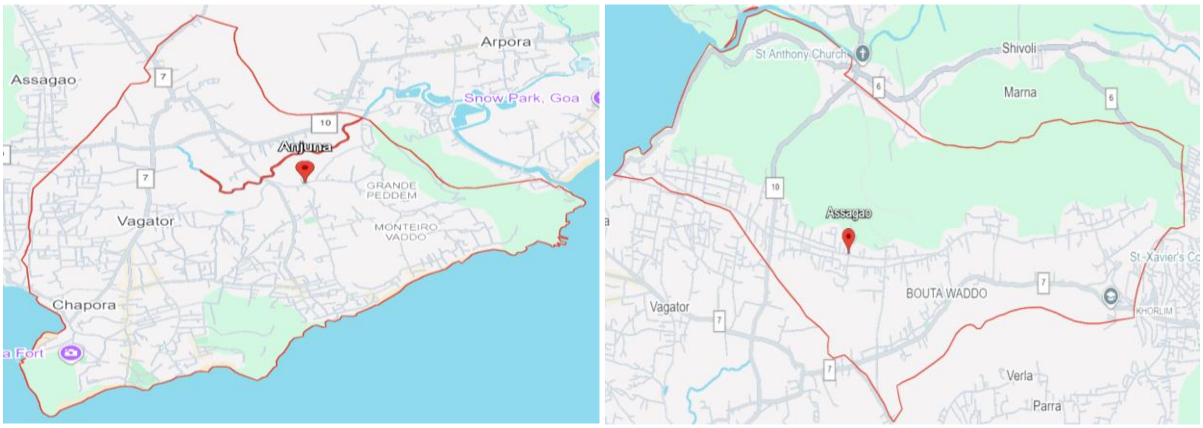

Figure 4.a. Area Under Anjuna Police station

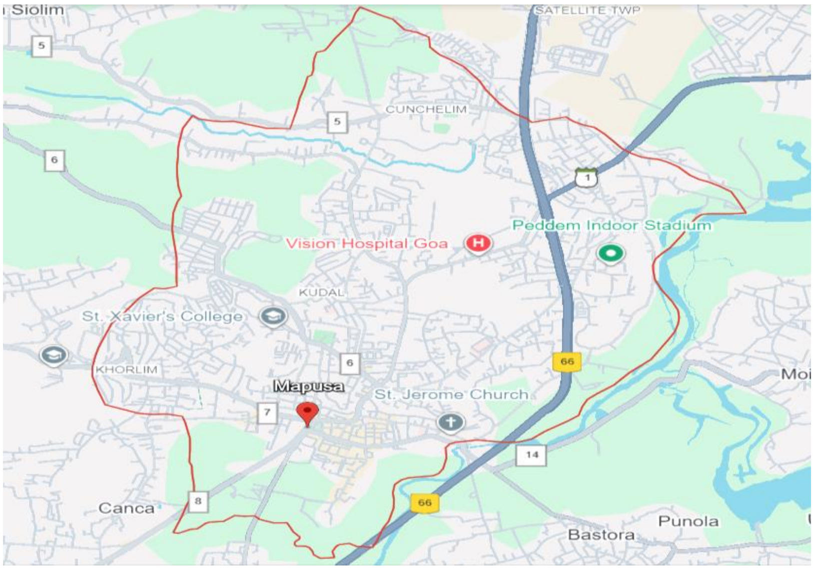

Figure 4.b. Area Under Mapusa Police station

Figure 4.c. Area Under Colvale Police station

The most frequently reported crimes were Assault(Hurt), Theft(Auto Theft), Drug-related(NDPS) Figure(Figure 4.1.2) . Notably, Drug-related(NDPS) and Assault(Hurt) (Figure 4.1.3) crimes were more common in tourist-heavy areas like Anjuna, with seasonal increases aligning with heightened tourism activity. And followed by Mapusa Theft (Auto theft) (Figure 4.1.4)

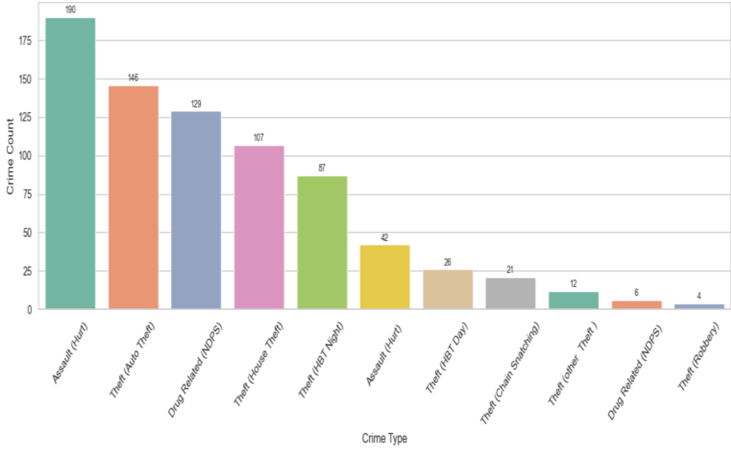

Figure 4.1.1 Most Common Crimes

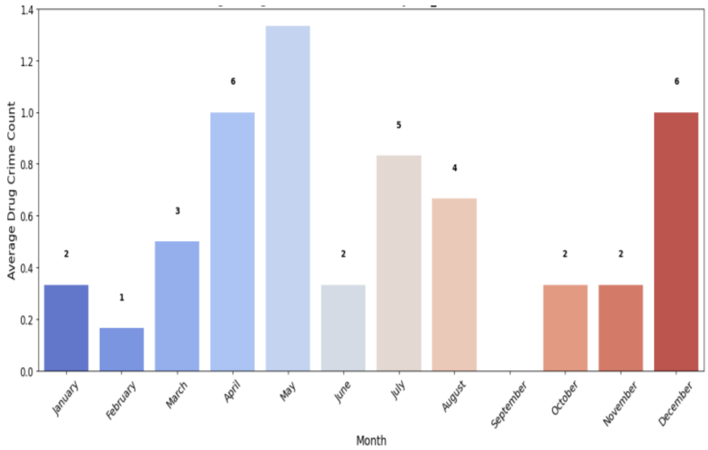

Figure 4.1.2 Drug cases in Anjuna

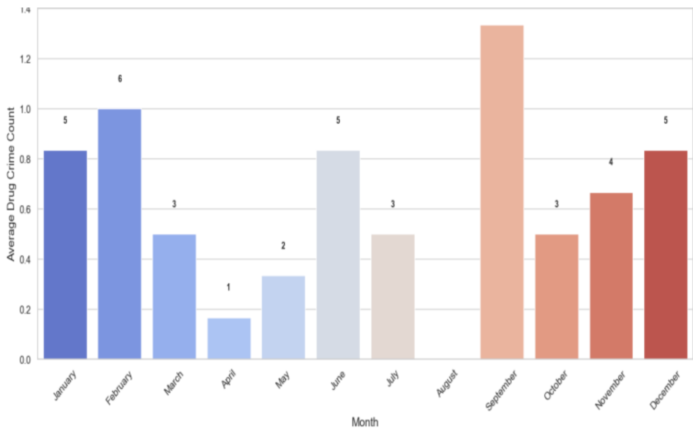

Figure 4.1.3 Assault (Hurt)cases in Anjuna

Followed by Theft (Auto theft) cases that is under Mapusa_Ps that is 92 cases.

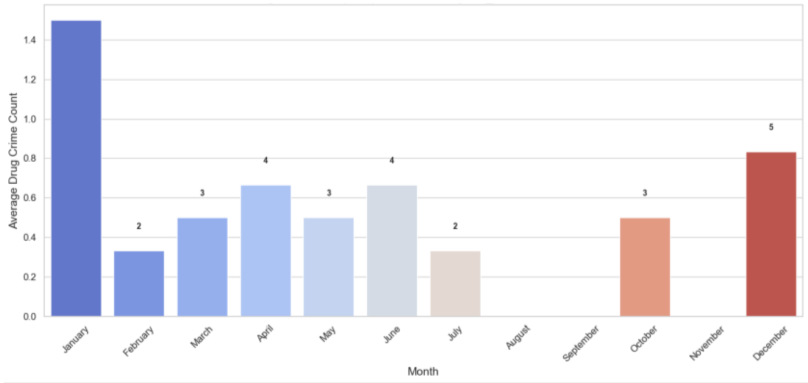

Figure 4.1.4 Average Theft (auto theft)

Most crimes took place in InHouse. Although residentialArea and LocalAreas also reported significant crime, they exhibited a more consistent pattern throughout the year. (Figure 4.1.5, Table 4.iii))

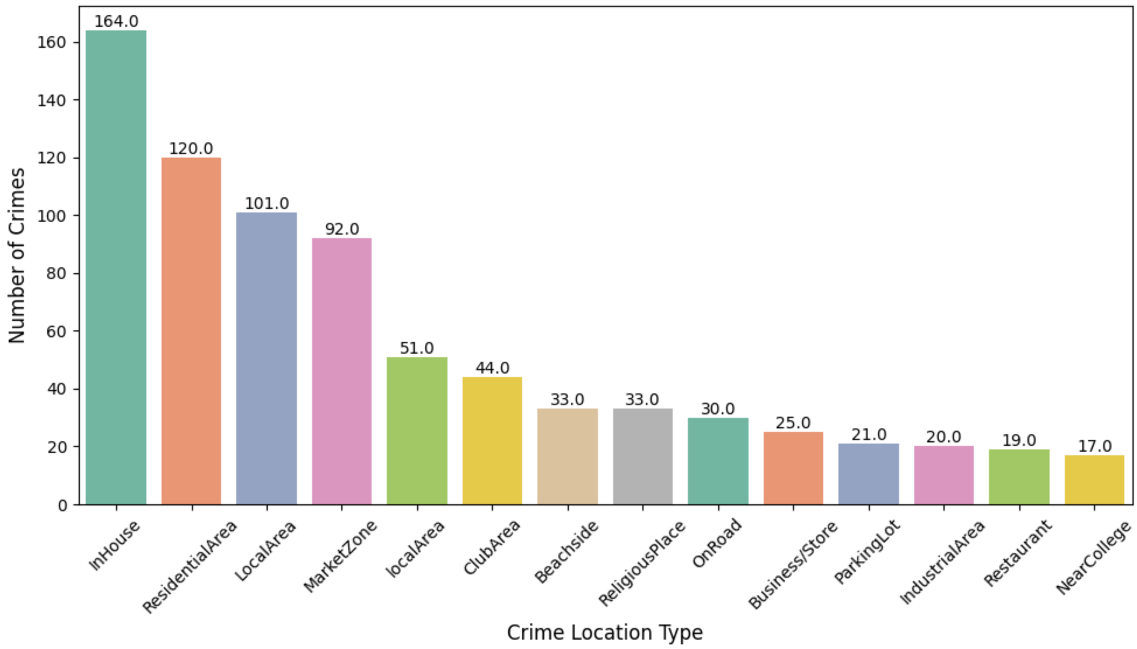

Figure 4.1.5 Total Crimes in all 3 areas

InHouse	**ResidentialArea**	**localArea**
1. Theft (House Theft) (46 cases) 2. Assault (Hurt) (37 cases) 3. Theft (HBT Night) (36 cases)	1. Assault (Hurt) (31 cases) 2. Theft (Auto Theft) (30 cases) 3. Drug Related (NDPS) (15 cases)	1. Assault (Hurt) (22 cases) 2. Drug Related (NDPS) (18 cases) 3. Theft (Auto Theft) (6 cases)

Table 4.iii) Interpretation of crime the top 3

Monthly crime rates over the years. We see January, February, March having the most amount of cases followed by July, April having high crime rates.

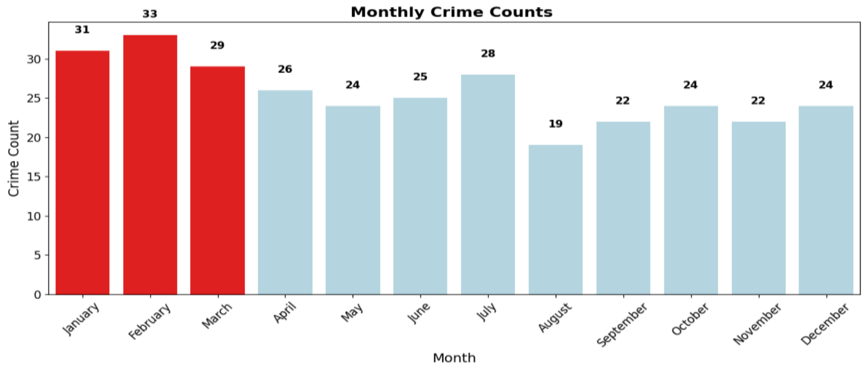

Figure 4.1.6 Total Crimes Monthly in all 3 areas

Analysing crime occurrences over time revealed distinct seasonal variations in **Mapusa and Anjuna(2019-2024)**, with a marked increase in crimes during Festive seasons(Ganesh Chaturthi, Diwali, Christmas, New Year). Especially during the New years in January and the Chaturthi month that is September we see a high crime rate than the other two festivals (Figure 4.1.7 & Figure 4.1.8)

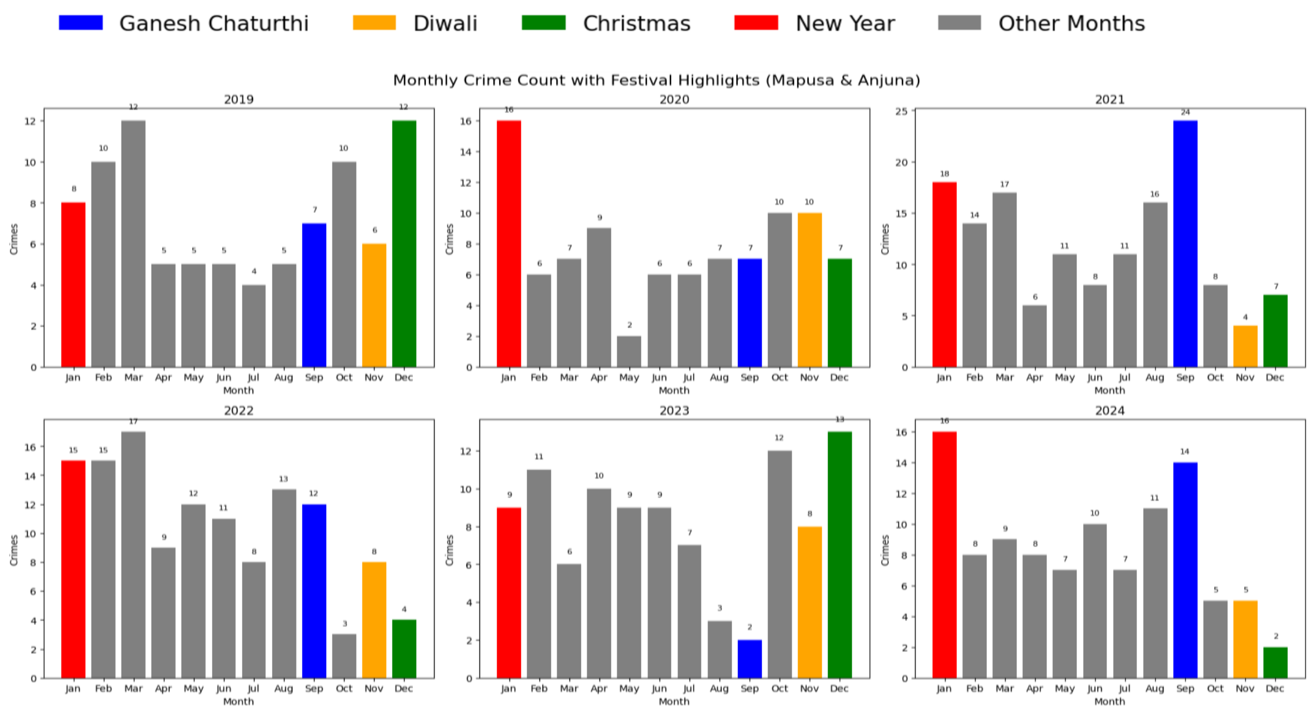

Figure 4.1.7 Monthly crime Count in Mapusa & Anjuna (Seasonal/Festivals)

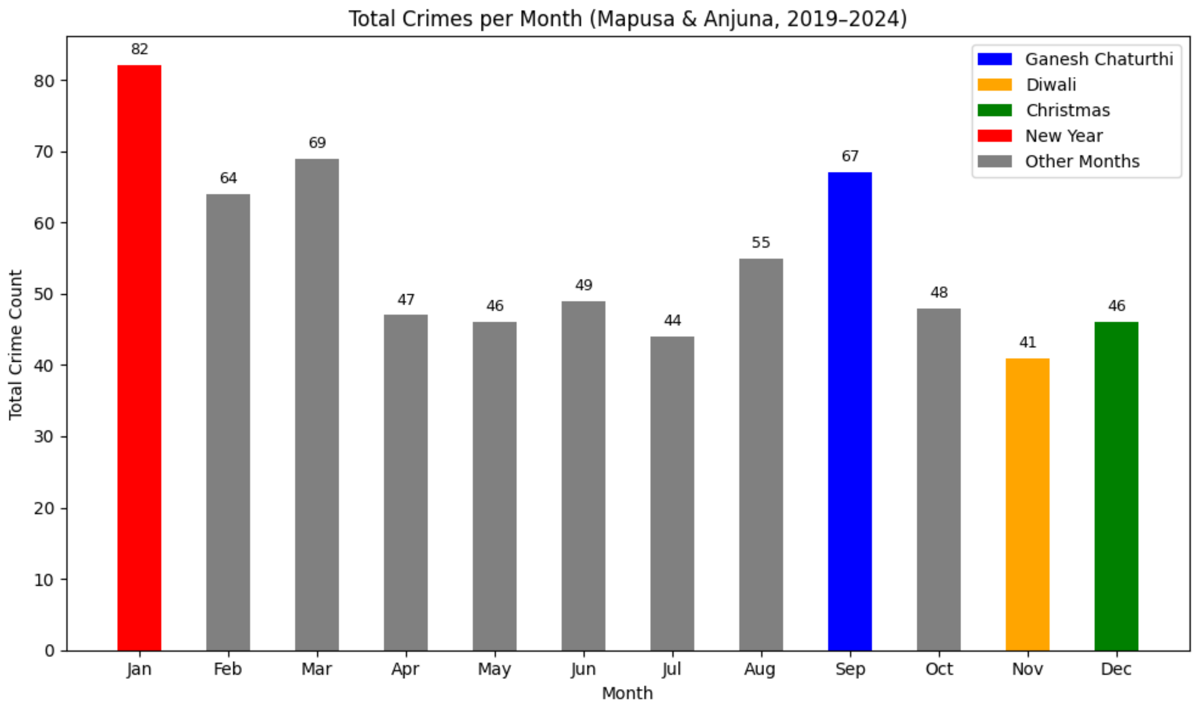

Figure 4.1.8 Monthly crime Count in Mapusa & Anjuna Average of all Months (Seasonal/Festivals)

Temporal Trends: Analysing crime occurrences over time revealed distinct seasonal variations in **Colvale(2021-2024)**, theres is no much increase in crimes during Festive seasons(Ganesh Chaturthi, Diwali, Christmas, New Year) in Colvale as compared to Mapusa and Anjuna. (Figure 4.1.9 & Figure 4.2.0)

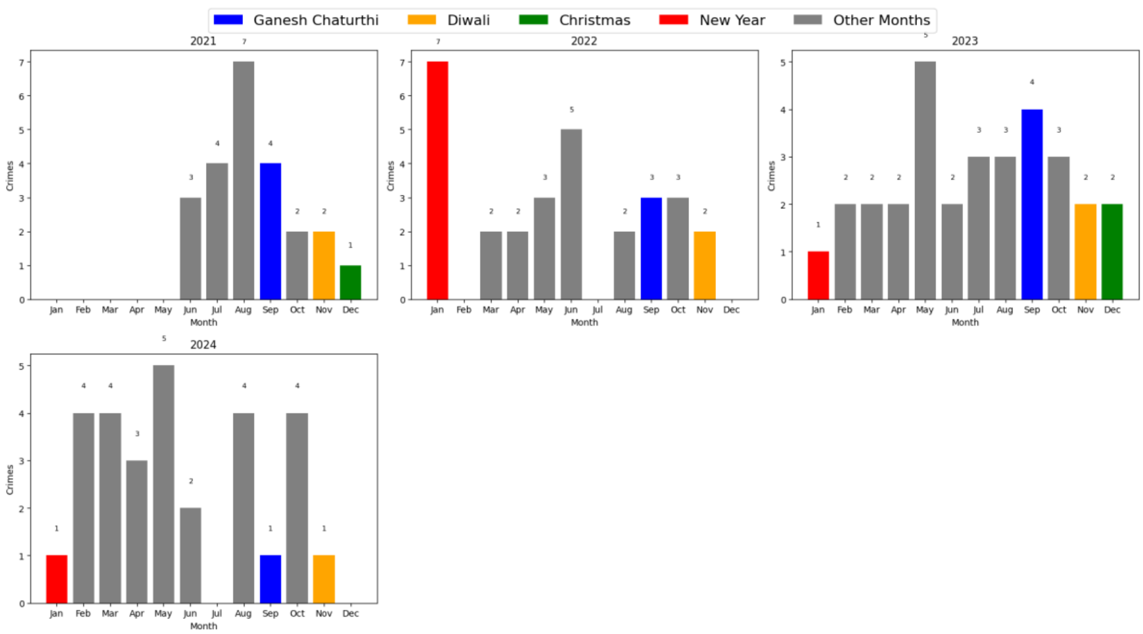

Figure 4.1.9 Monthly crime Count in Colvale

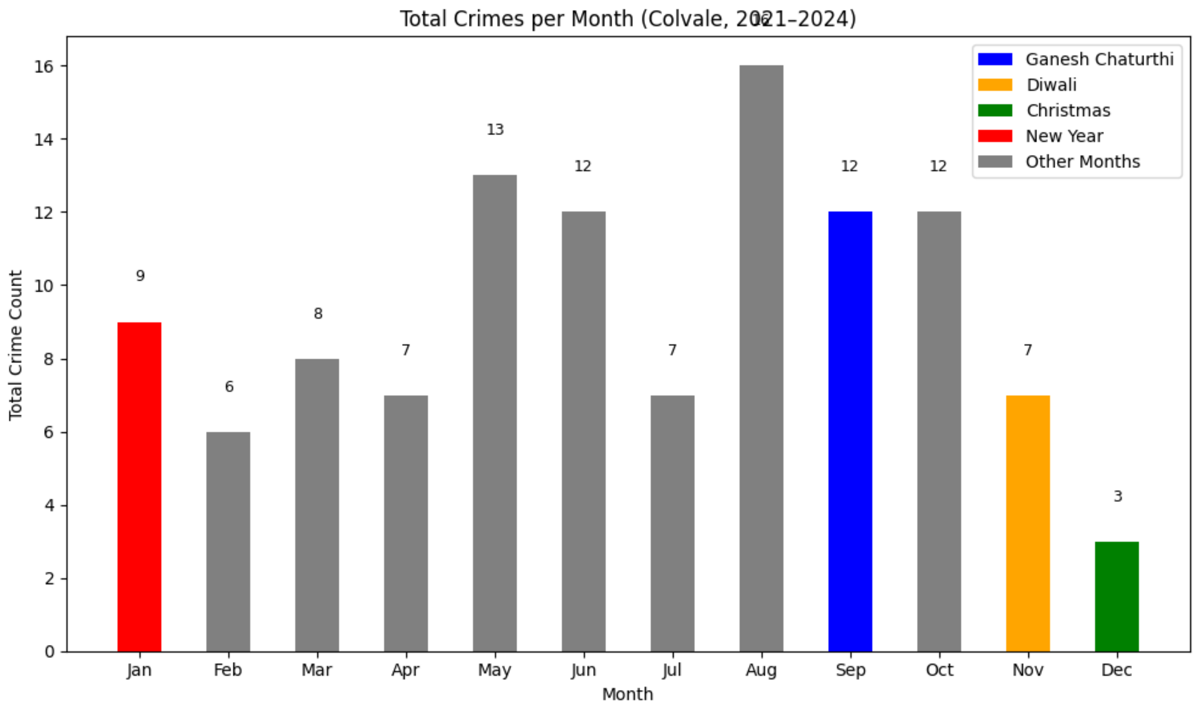

Figure 4.2.0 Monthly crime Count in Colvale

Weekday trends showed higher crime frequencies on weekdays than weekends.

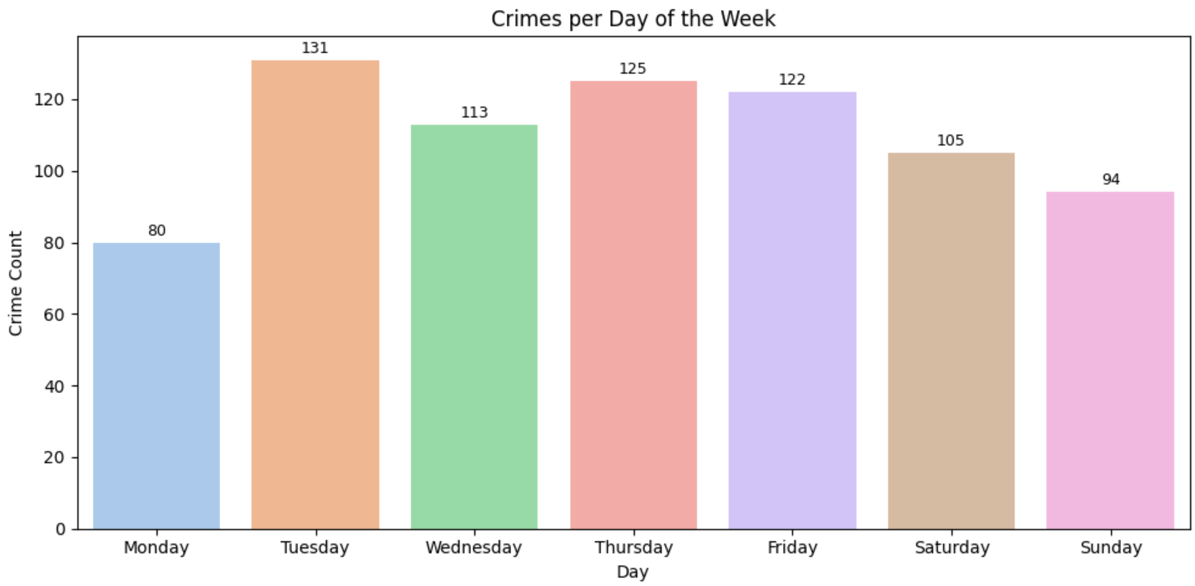

Figure 4.2.1 crimes per days of week

Crime Hotspot Detection and Spatial Analysis. The clustering algorithms, specifically K-Means, offered valuable insights into high-density crime zones. High-Risk Areas: Crime hotspots were identified in areas surrounding Anjuna, where the concentration of crimes were Assult(Hurt) with 107 case and drug-related(NDPS) with 89 cases, was the highest.

Recommendations for Police Deployment

The system generated recommendations for police deployment by examining crime data and forecasting areas with high activity levels. It was suggested that patrols focus more on crime hotspots identified through clustering algorithms, particularly in Anjuna and Mapusa during peak tourist seasons, where there is a noted increase in drug-related crimes. Crime Type-Specific Deployment: Deployment strategies should be customized according to the type of crime. For instance, Anjuna should receive additional resources during the tourist season to manage the surge in drug-related offenses. Similarly, Mapusa should be closely observed for thefts, especially in residential and local areas, where these crimes are more prevalent.

Dashboard:

Figure 4.2.2 Crime Analytics & Police Deployment Dashboard

Figure 4.2.3 Crime Analytics & Police Deployment Dashboard Selecting Crime type

Output for the given crime type:

1. Displays the Top reporting police station.
2. Most common crime location.
3. Predicted crime location using Knn
4. Police deployment recommendations with respect to each of the police stations
5. Displays Crime Hotspot mapping for the selected Crime type

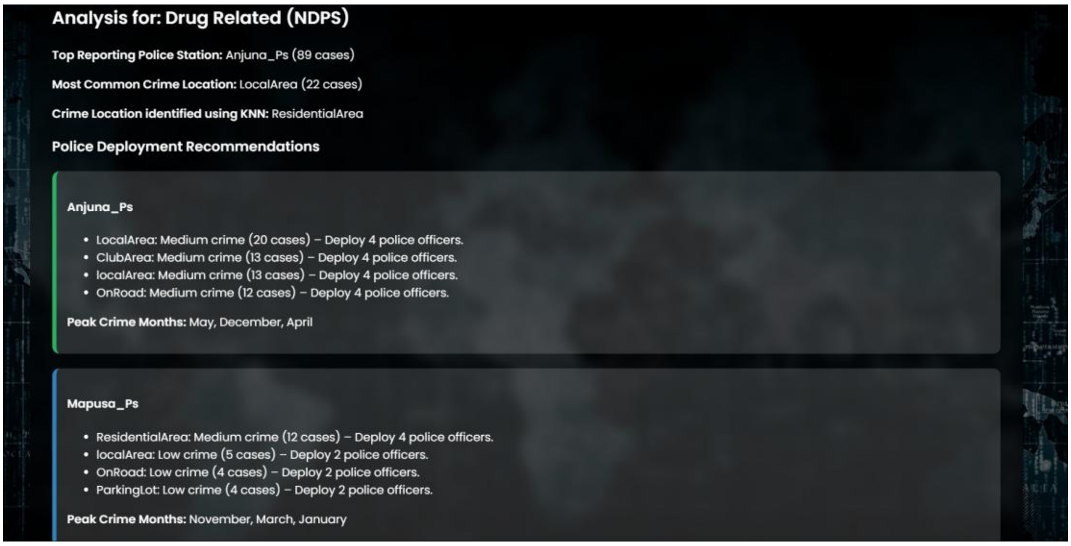

Figure 4.2.4 Crime analysis for the selected Crime type

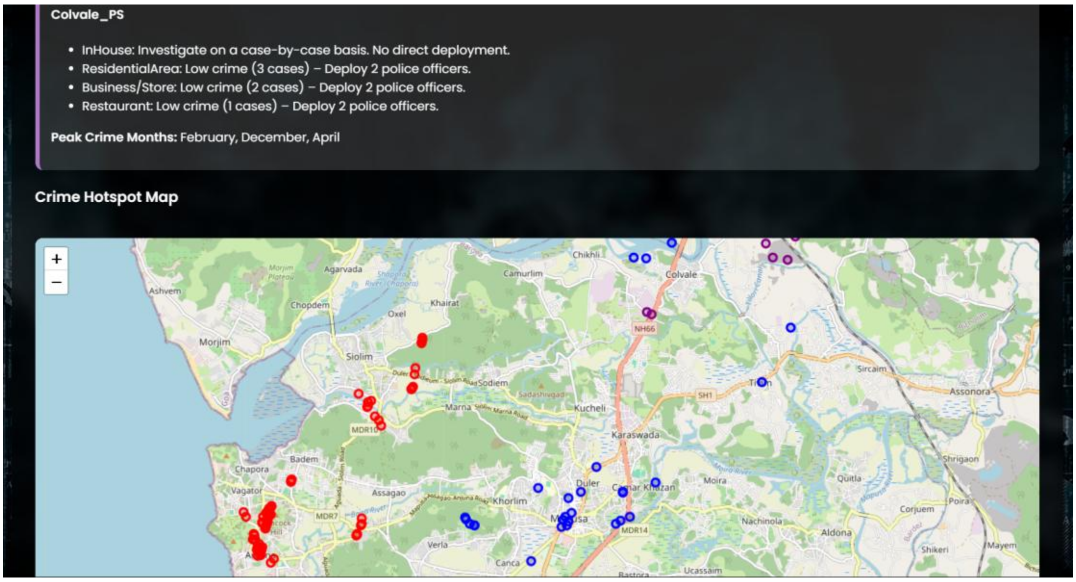

Figure 4.2.5 Crime Hotspot mapping for the selected Crime type

Conclusion

This project successfully showcased the use of machine learning and spatial analysis techniques on crime data to enhance police resource allocation and proactive crime prevention. The K-Nearest Neighbours (KNN) model effectively predicted crime location types with a DB score of 78.8, supporting classification-based strategies. Clustering algorithms also helped pinpoint critical crime hotspots, aiding in targeted deployment efforts. Integrating Geographic Information Systems (GIS) with machine learning facilitated a data-driven approach to improve public safety.

Future Scope

The analysis relied on historical crime data, which may have reporting biases or gaps. Future systems should incorporate real-time crime data for more accurate and timely predictions. The existing model used limited features such as spatial data and crime types. Including additional variables like socioeconomic factors, demographic patterns, and real-time events (e.g., festivals, rallies) could significantly enhance prediction accuracy. There is potential for experimenting with more advanced machine learning models and ensemble methods to further improve performance. A real-time alert system can be developed to notify authorities when a hotspot shows signs of unusual activity, enabling quicker response times."

References

1. Agarwal, J. (2013). Crime analysis using K-Means clustering. International Journal of Computer Applications, 83. http://dx.doi.org/10.5120/14433-2579
2. Alqahtani, A., Garima, A., & Alaiad, A. (2019). Crime analysis in Chicago City. In 2019 10th International Conference on Information and Communication Systems (ICICS 2019) (pp. 166–172). https://doi.org/10.1109/IACS.2019.8809142
3. Anselin, L., Gorr, W., & Tita, G. (2000). Spatial analyses of crime. https://www.researchgate.net/publication/235217015
4. Bernasco, W., & Elffers, H. (2010). Statistical analysis of spatial crime data. In A. Piquero & D. Weisburd (Eds.), Handbook of Quantitative Criminology (pp. 699–724). Springer New York. https://doi.org/10.1007/978-0-387-77650-7_33
5. Data & Society. (2023). Ethics of AI surveillance: A research brief. https://datasociety.net/library/
6. ESRI. (2022). GIS and crime mapping: Technical guide. https://www.esri.com/en-us/industries/public-safety/overview
7. Fredrick David, H. B., & Suruliandi, A. (2017). Survey on crime analysis and prediction using data mining techniques. ICTACT Journal on Soft Computing, 7(3), 1459–1466. https://doi.org/10.21917/ijsc.2017.0202
8. Gahalot, A., Uprant, Dhiman, S., & Chouhan, L. (2020). Crime prediction and analysis. In 2nd International Conference on Data, Engineering and Applications (IDEA 2020). https://doi.org/10.1109/IDEA49133.2020.9170731
9. Hassani, H., Huang, X., Silva, E. S., & Ghodsi, M. (2016). A review of data mining applications in crime. Statistical Analysis and Data Mining, 9(3), 139–154. https://doi.org/10.1002/sam.11312
10. Kedia, P. (2016). Crime mapping and analysis using GIS. https://doi.org/10.13140/RG.2.2.11064.14081
11. Koren, G. (2021). Ten Commandments for patient-centred treatment. Clinical and Translational Science, 15(2), e1046. https://doi.org/10.1111/cts.13046
12. Kounadi, O., Ristea, A., Araujo, A., & Leitner, M. (2020). A systematic review on spatial crime forecasting. Crime Science, 9(1). https://doi.org/10.1186/s40163-020-00116-7

13. Kukreja, G., Mav, S., Vaswani, A., Pathak, P., Bhatia, M., & Hod, D. (2018). Crime data analysis. International Journal of Creative Research Thoughts (IJCRT), 6(1). www.ijcrt.org
14. Mandalapu, V., Elluri, L., Vyas, P., & Roy, N. (2023). Crime prediction using machine learning and deep learning: A systematic review and future directions. IEEE Access, 11, 60153–60170. https://doi.org/10.1109/ACCESS.2023.3286344
15. National Institute of Justice. (2023). Data analytics for public safety: A review of government and NGO initiatives (2020–2023). U.S. Department of Justice. https://nij.ojp.gov/library/publications/data-analytics-public-safety
16. Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2015). Predictive policing: The role of crime forecasting in law enforcement operations. Crime Science, 4, 1. https://doi.org/10.1186/s40163-015-0041-6
17. Pratibha, D., Ritisha, D., Pooja, R., & Jasmine, K. S. (n.d.). Crime data analysis and prediction.
18. Raji, I., & Sholademi, D. B. (2023). Predictive policing: The role of AI in crime prevention. International Journal of Computer Applications Technology and Research, 13(10), 248–252. https://ijcat.com/archieve/volume13/issue10/ijcatr13101006.pdf
19. ResearchGate. (n.d.). How to categorize crime data into HIGH, MEDIUM, and LOW. Retrieved April 14, 2025, from https://www.researchgate.net/post/How_to_categorize_crime_data_into_HIGH_MEDIUM_and_LOW
20. Reid, S. E., Tita, G., & Valasik, M. (2011). The mapping and spatial analysis of crime. In Criminology. Oxford University Press. https://doi.org/10.1093/obo/9780195396607-0123
21. Ristea, A., & Leitner, M. (2020). Urban crime mapping and analysis using GIS. ISPRS International Journal of Geo-Information, 9(9). https://doi.org/10.3390/ijgi9090511
22. Salah, K., & Alcaraz Calero, J. M. (2011). Cloud computing for distributed university campus: A prototype using Google App Engine. IEEE Journal on Selected Areas in Communications, 29(10), 1986–1992. https://doi.org/10.1109/JSAC.2011.110122
23. Sieveneck, S., & Sutter, C. (2021). Predictive policing in the context of road
24. traffic safety: A systematic review and theoretical considerations. Transportation Research Interdisciplinary Perspectives, 11, 100429. https://doi.org/10.1016/j.trip.2021.100429

25. Smith, R. (2023). Predictive policing and artificial intelligence models. International Journal of Law and Technology, 18(2), 87–99. https://academic.oup.com/ijlit
26. Tamir, A., Watson, E., Willett, B., Hasan, Q., & Yuan, J.-S. (2021). Crime prediction and forecasting using machine learning algorithms. International Journal of Computer Science and Information Technology. https://www.researchgate.net/publication/355872171
27. Tyagi, D., & Sharma, S. (2020). An approach to crime data analysis: A systematic review. International Journal of Engineering Technologies and Management Research, 5(2), 67–74. https://doi.org/10.29121/ijetmr.v5.i2.2018.615
28. Ulla Khan, T. (2022). Crime analysis and prediction using machine learning. International Research Journal of Engineering and Technology (IRJET). www.irjet.net
29. United Nations Office on Drugs and Crime. (2021). Global smart policing: Urban crime pattern analysis. https://www.unodc.org/unodc/en/publications.html