Vivekanand Education Society's
Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai, Approved by AICTE & Recognised by Govt. of Maharashtra)
NAAC accredited with 'A' grade

Semester: VI Subject : AIDS - 1

Title of the Project:
Smart Travel Safety Recommendation

Domain: Machine Learning

Group Members:
Member 1 : Avan Shetty
Member 2: Rakshit Sharma
Member 3: Veydant Sharma

Professor Name: Dr. Ravita Mishra

Content

- Overview
- Introduction
- Problem Statement
- Objectives
- Literature Survey
- Proposed System
- Proposed Design (along with UML Diagrams)
- Implementation
- Conclusion
- References

Introduction to Project

In recent years, India has witnessed a notable rise in various criminal activities. For instance, the National Crime Records Bureau (NCRB) reported that the crime rate decreased from 487.8 per 100,000 people in 2020 to 445.9 in 2021 further it was reduced to 442.2 in 2023 . Additionally, the rapid adoption of digital transactions has positioned India as a global leader in digital payments, but it has also led to a surge in digital financial crimes.

The finance ministry reported that high-value cyber fraud cases have surged, costing millions. These statistics underscore the pressing need for accurate and accessible information on crime rates across different regions to ensure the safety of travelers

Problem Statement

The NCRB provides detailed crime data, but it's often too complex for the general public to understand. This makes it harder for travelers to assess safety, increasing their risk. With rising cybercrimes targeting tourists unfamiliar with local systems, there's a growing need for a platform that simplifies both physical and digital safety information.

Addressing this issue requires developing a platform that aggregates, analyzes, and presents crime data in an accessible manner, empowering travelers to make informed and safe choices. The project aims to **predict crime-prone areas** using crime datasets and classification algorithms, ensuring safer travel recommendations and information about the crime rate

Objectives of the project

•To analyze crime trends and classify regions based on safety levels.

•To develop a machine learning model that predicts crime-prone areas.

•To implement classification algorithms such as KNN**, Decision Trees, clustering and Random Forest** to assess model performance regarding safe or unsafe and also regression to predict the crime rate at that particular place

•To provide **visual insights** through heatmaps and crime density plots.

•To build an interactive **dashboard using Streamlit** for real-time analysis.

Literature Survey

Sr.No	Title	Author	Publish Date/Year	Description
1	Crime Analyses Using Data Analytics	Thanu Dayara, Manukau Institute of Technology, New Zealand Fadi Thabtah, ASDTests, Auckland, New Zealand Hussein Abdel-Jaber, Arab Open University, Saudi Arabia* Susan Zeidan, Zayed University, UAE	2022	This study uses machine learning to predict arrest outcomes for serious crimes in Chicago based on time, location, and crime type. High-accuracy models help optimize law enforcement resources and crime prevention. The classification models derived in the paper achieved an accuracy of around **80.22%** for predicting arrest status. After applying the **(SMOTE)** to balance the dataset, the **Decision Tree (C4.5) algorithm** achieved the highest predictive accuracy of **63.50%** .
2.	Crime Prediction Model using Three Classification Techniques: Random Forest, Logistic Regression, and LightGBM	Abdulrahman Alsubayhin, Muhammad Sher Ramzan, Bander Alzahrani King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia	2024	This study compares machine learning models for crime prediction, with LightGBM achieving the highest accuracy, followed by Random Forest and Logistic Regression. These models help predict crime likelihood and improve law enforcement strategies.

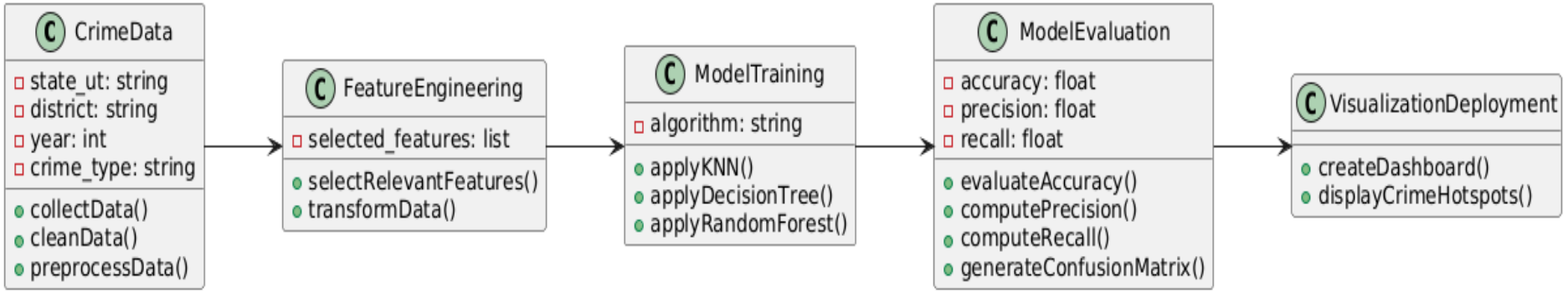
Proposed System

- **Data Collection & Cleaning** – Preprocessing crime dataset of India.
- **Feature Engineering** – Selecting relevant features like crime type, location, time , murder, attempt to murder, culpable homicide not amounting to murder, rape, custodial rape, other rape etc
- **Model Training** – Applying KNN, Decision Tree, and Random Forest for classification.
- **Model Evaluation** – Using accuracy, precision, recall, and confusion matrix.
- **Visualization & Deployment** – Displaying crime-prone areas via an interactive dashboard (Streamlit)

Dataset Overview

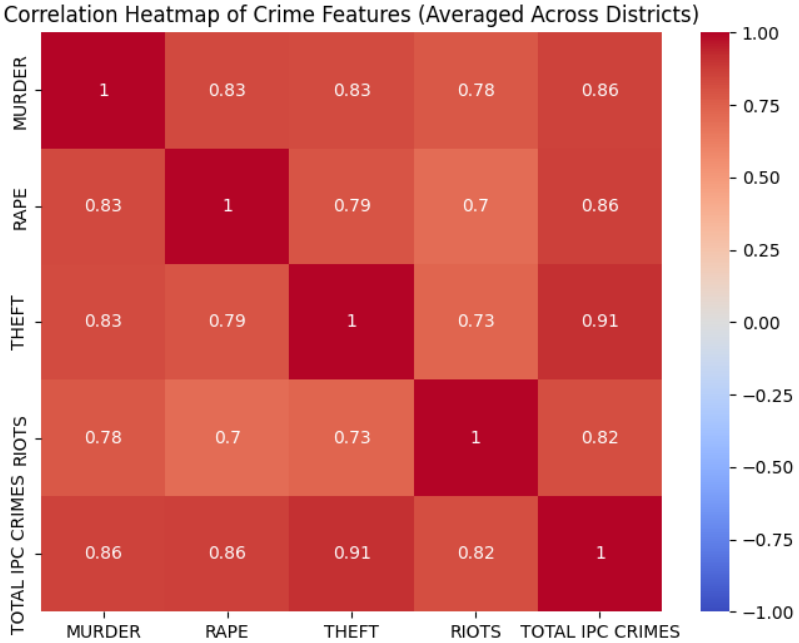
	STATE/UT	DISTRICT	YEAR	MURDER	ATTEMPT TO MURDER	CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	RAPE	CUSTODIAL RAPE	OTHER RAPE	KIDNAPPING & ABDUCTION	...	HURT/GREVIOUS HURT	DOWRY DEATHS	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	INSULT TO MODESTY OF WOMEN	CRUELTY BY HUSBAND OR HIS RELATIVES	IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES	CAUSING DEATH BY NEGLIGENCE	OTHER IPC CRIMES	TOT I CRIM
0	ANDHRA PRADESH	ADILABAD	2001	101	60	17	50	0	50	46	...	1131	16	149	34	175	0	181	1518	41
1	ANDHRA PRADESH	ANANTAPUR	2001	151	125	1	23	0	23	53	...	1543	7	118	24	154	0	270	754	41
2	ANDHRA PRADESH	CHITTOOR	2001	101	57	2	27	0	27	59	...	2088	14	112	83	186	0	404	1262	58
3	ANDHRA PRADESH	CUDDAPAH	2001	80	53	1	20	0	20	25	...	795	17	126	38	57	0	233	1181	31
4	ANDHRA PRADESH	EAST GODAVARI	2001	82	67	1	23	0	23	49	...	1244	12	109	58	247	0	431	2313	65

Proposed Design



Implementation

1. Crime Heatmap – Visualizing crime intensity across different regions.



The **high correlations** indicate that crime in a region is not usually isolated but rather an interlinked pattern. This suggests that predicting crime risk for travelers based on a few major categories (like theft, murder, and rape) can provide a reliable safety assessment for a given location

Implementation

2. Random Forest Model – Showing classification results for safe vs. unsafe areas.

Setting the threshold and the features highly correlated were selected

```
features = ['MURDER', 'RAPE', 'THEFT', 'RIOTS']
X = df[features]


median_crime = df['TOTAL IPC CRIMES'].median()
df['SAFETY_LABEL'] = (df['TOTAL IPC CRIMES'] > median_crime).astype(int)
y = df['SAFETY_LABEL']
```

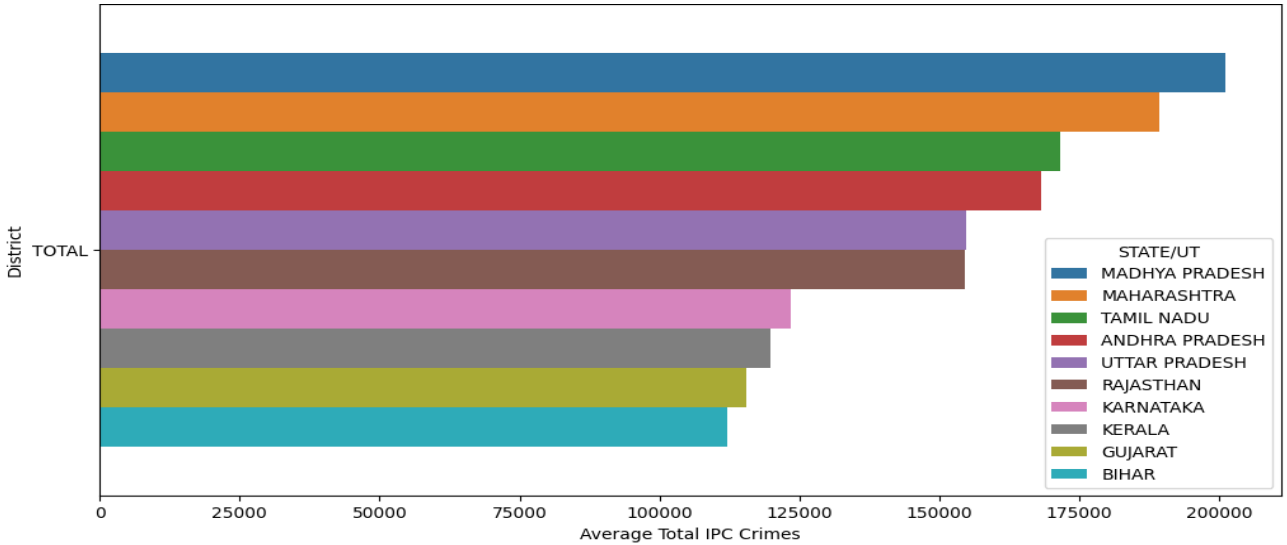
Metrics given the model

```
Accuracy: 0.8991130820399114
              precision    recall  f1-score   support

           0       0.90      0.90      0.90       903
           1       0.90      0.90      0.90       901

    accuracy                           0.90      1804
   macro avg       0.90      0.90      0.90      1804
weighted avg       0.90      0.90      0.90      1804
```

The **f1-score of 0.90** confirms that the model maintains a good balance between precision and recall. The **support values (903 for safe, 901 for unsafe)** suggest an even distribution of the dataset, ensuring that the model does not favor one class over the other

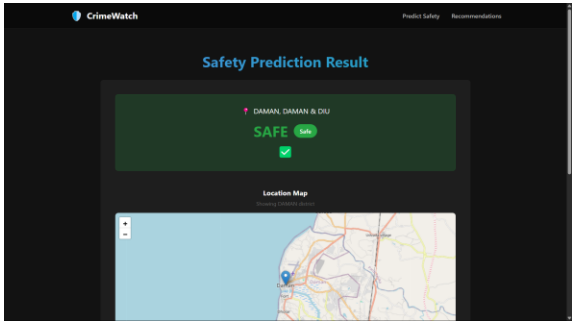
Implementation

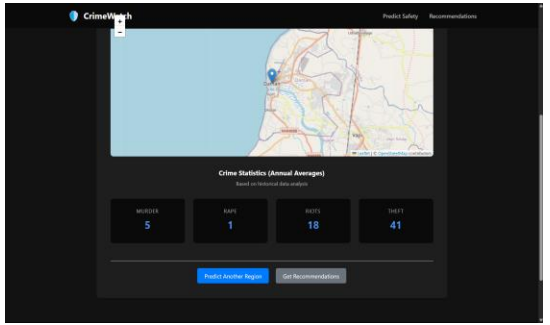
3. Geographical Crime Map – Displaying safety predictions on an interactive map

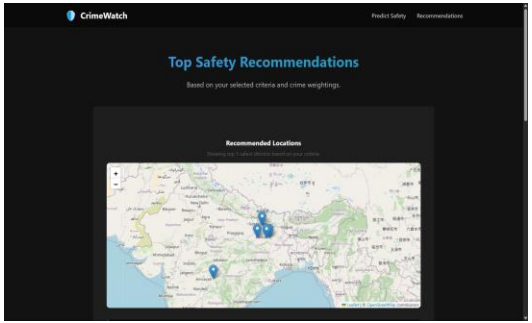


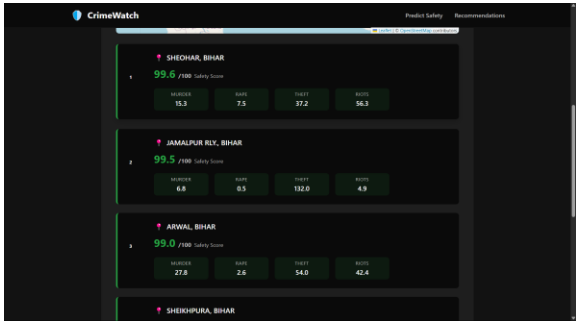
This analysis helps identify high-risk states, allowing for better crime awareness and planning. Travelers can use this data-driven insight to assess safety before visiting specific locations

Implementation









Conclusion

Our model classifies crime data into safe and unsafe categories using KNN and Random Forest, helping travelers make informed decisions based on regional crime patterns. By leveraging historical crime data, it offers a data-driven approach to assessing safety. While effective, future improvements could introduce finer safety tiers (eg Safety-1, Safety-2 or Unsafe-1, Unsafe-2) for more precise risk evaluation, further enhancing travel awareness and making it more scalable for travellers.

References

1.https://www.researchgate.net/publication/360166504_Crime_Analyses_Using_Data_Analytics/link/635febbc96e83c26eb6e5ad3/download

2. https://thesai.org/Downloads/Volume15No1/Paper_23-Crime_Prediction_Model_using_Three_Classification_Techniques.pdf?utm_source=chatgpt.com

3. Crime in India

4. Crime Statistics - Dataset - India Data portal