www.ijarp.com ISSN 2456-9992 Page: 01-13

International Journal Advanced Research Publications

Research Article

Volume: 02
Issue: 05

SAFEROUTE - AN ML BASED SAFETY PREDICTION AND THREAT ANALYSIS SYSTEM FOR WOMEN TOURISTS

***Harsh Gupta, Aarya Mhetre, Aditi Hiray, Shrija Kahadane, Dr. Rohini Palve**

Department of Computer Engineering Terna Engineering College, Nerul, Navi Mumbai, India.

Article Received: 09 April 2026, Article Revised: 29 April 2026, Published on: 19 May 2026

***Corresponding Author: Harsh Gupta**

Department of Computer Engineering Terna Engineering College, Nerul, Navi Mumbai, India.

DOI: https://doi-doi.org/101555/ijarp.9391

ABSTRACT

Women tourists face heightened safety risks in unfa- miliar environments due to the lack of real-time, location-aware safety intelligence. *SafeRoute* is a web-based platform designed to address this gap by integrating machine learning, geospatial visualization, and emergency response mechanisms to enhance the safety of women travelers. The system enables citizens to report safety incidents anonymously or with identification and employs an XGBoost-based classifier to predict crime-prone areas using historical incident data obtained from the National Crime Records Bureau (NCRB) [1]. A pre-trained AI model detects and filters fake or misleading reports to ensure data integrity. Crime hotspots are dynamically visualized through an intensity- weighted heatmap rendered using the Leaflet.js mapping library and its Leaflet.heat plugin, where each district's geographical coordinates are mapped against total crime volume to produce color-coded spatial overlays enabling tourists to identify high-risk zones and plan safer routes. A role-based dashboard architecture serves tourists, safety officers, and administrators with tailored interfaces for reporting, monitoring, and response. An integrated SOS module allows tourists to trigger real-time emergency alerts with automatic location sharing to nearby safety officers. Evaluation of the prediction model yielded an accuracy of 87.46% and an AUC of 0.9572, with macro-averaged precision, recall, and F1-score all at 0.87, demonstrating reliable performance across balanced class distributions. *SafeRoute* offers a scalable, data- driven solution to improve situational awareness and emergency responsiveness for women tourists.

***INDEX TERMS*:** Women Tourist Safety, Crime Prediction, Ma- chine Learning, XGBoost,

Heatmap Visualization, Leaflet.js, SOS Emergency Feature, Pre-trained AI Model, Fake Report Detection, Predictive Analytics

I. INTRODUCTION

Tourism is one of the fastest-growing sectors globally, yet women tourists continue to face disproportionate safety risks in unfamiliar destinations. Incidents of harassment, assault, and theft targeting women travelers are frequently underre- ported due to language barriers, social stigma, and the absence of accessible reporting mechanisms. Traditional safety infras- tructure fixed police stations, reactive complaint systems, and printed safety advisories fails to meet the dynamic, location- sensitive needs of modern travelers. The absence of predictive intelligence and real-time threat assessment further limits the ability of authorities and tourists to make informed safety decisions.

Existing research on crime prediction has largely focused on static, region-specific datasets with limited applicability to tourist populations [2]–[4]. Systems that integrate AI form women's safety have typically targeted residents rather than transient visitors, and rarely incorporate anonymous reporting, spatial hotspot visualization, or emergency response in a single unified platform [5], [8]. Furthermore, the proliferation of false or misleading incident reports poses a data quality challenge that reduces the reliability of crime prediction models [7], [9], [11].

To address these limitations, this paper presents *SafeRoute*, an intelligent web-based platform designed specifically for the safety of women tourists. The system contributes the following:

An ML-based crime prediction engine using **XGBoost**

trained on historical NCRB crime data;

Dynamic crime hotspot visualization using an **intensity- weighted heatmap** rendered via Leaflet.js and the Leaflet.heat plugin, where district-level crime counts serve as spatial weights to generate color-coded overlays red indicating high-risk zones, green indicating safer areas;

Anonymous and registered incident reporting with a **pre- trained AI model** for fake report detection;

An integrated **SOS module** that enables real-time emer- gency alerts with automatic GPS-based location sharing to assigned safety officers; and A role-based multi-dashboard architecture for tourists, safety officers, and administrators.

The remainder of this paper is organized as follows. Sec- tion II reviews related literature. Section III describes the limitations of existing systems and the problem statement. Section

IV presents the proposed system methodology. Sec- tion V details the system architecture and implementation. Section VI discusses results. Section VII addresses challenges and limitations, and Section VIII concludes the paper.

II. Literature Review

Shalini et al. [2] introduced a crime prediction technique using the Random Forest model. Their system analyzed histor- ical crime data to classify crime-prone areas and crime types, handled missing data efficiently, and visualized trends by crime type and region. Random Forest outperformed Decision Tree models in pattern recognition and prediction accuracy. However, the study was conducted in offline experimental settings and lacked real-time integration, interactive dash- boards, and fake-report detection all of which are essential for a tourist-facing safety system where data freshness and reliability are critical.

Patel and Zala [3] applied supervised regression models to analyze and predict crime against women in India. LASSO regression achieved the best performance with R^2 of 0.99 and low prediction errors. While effective for offline predictive analytics, the system lacked real-time user input, interactive dashboards, and anonymous reporting features necessary to support women tourists who may be unfamiliar with local reporting channels.

Mwaniki et al. [4] evaluated Decision Tree, Random Forest, and hybrid AdaBoost algorithms for crime type prediction using historical data including location and time. The hybrid model achieved the highest accuracy of 94.4%. However, the study was limited to offline experiments with no real-time system implementation or web interface, making it unsuitable for deployment in a tourist safety context that demands live prediction and map-based visualization.

Subha et al. [5] proposed an AI-based integrated women safety system incorporating real-time monitoring, a mobile application, and a physical emergency unit. The system used AI-driven video anomaly detection and GPS-based SOS alerts to police and emergency contacts. Although the system sup- ported real-time deployment and hardware-based triggers, it was designed for residents with physical devices—not for tourists who require a lightweight, web-accessible platform with no hardware dependency.

Kavala et al. [6] described a crime analysis and prediction system using Multi-Layer Perceptron (MLP) neural networks and Random Forest models, achieving 93% accuracy with MLP. Visualizations were provided via Google Data Studio. The system was limited to offline predictive modeling and lacked real-time complaint integration, multi-role dashboards, and fake-report detection gaps that *SafeRoute* directly ad- dresses for tourist safety use cases.

Srivastava and Singh [7] proposed a spam detection frame- work using TF-IDF, N-grams, Naive Bayes, and Support Vector Machine models for classifying malicious content. The approach demonstrated high classification accuracy in lab- based simulations. This work informed the need for automated false-report detection in citizen-driven crime reporting systems, a challenge addressed in *SafeRoute* through a pre-trained AI model tailored to incident report validation.

Surana et al. [8] developed a chatbot-based crime regis- tration and awareness system using a custom Named Entity Recognition model, achieving complaint classification accu- racy of 92.6%. Although the system improved accessibility of complaint filing, it was limited to Indian legal contexts and lacked hotspot visualization, severity analysis, or SOS capabilities all of which are essential for serving tourists across diverse geographic locations.

Gautam et al. [9] developed a chatbot-based FIR registration system allowing citizens to file complaints without visiting police stations. The system improved reporting accessibility but lacked integration with police databases, crime prediction features, hotspot visualization, and fake-report validation. For women tourists who may not know local police procedures, a system that goes beyond complaint filing to provide predictive and emergency features is required.

Sharma and Dronavalli [10] presented a crime data analysis and visualization system using Tableau and GIS-based geocod- ing, mapping crime trends with socioeconomic factors to identify high-risk zones. The system supported data visualiza- tion and spatial analysis but included no predictive modeling or real-time processing. *SafeRoute* builds on this concept of spatial crime representation using intensity-weighted heatmaps via Leaflet.js and Leaflet.heat, extending it with live data and tourist-specific route safety guidance.

Jain et al. [11] proposed a machine learning-based fake news detection system using Naive Bayes, SVM, and NLP techniques, achieving 93.5% accuracy. The system's mod- ular architecture for content authentication highlighted the importance of automated verification in crowd-sourced data platforms. This principle directly motivates the inclusion of a pre-trained AI model for fake report detection in *SafeRoute*, ensuring that crowd-reported tourist safety incidents are reli- able before influencing predictions or alerts.

III. METHODOLOGY

The proposed system, *SafeRoute*, is designed to predict regional safety levels and assist in crime reporting through a combination of machine learning models and user-interactive modules. The methodology is divided into three major compo- nents: data processing and safety prediction, crime reporting, and user interaction modules.

A. Data Collection and Preprocessing

The dataset used in this study is obtained from the National Crime Records Bureau (NCRB), containing district-wise crime records against women for the years 2019 to 2022. The dataset includes features such as rape, dowry deaths, acid attacks, cruelty by husband or relatives, trafficking, and other relevant crime categories.

Data preprocessing is performed to ensure quality and consistency. This includes:

- Handling missing or inconsistent values
- Normalizing numerical features
- Aggregating crime counts district-wise
- Computing a severity ratio to represent the intensity of crimes in each region

The processed dataset is then labeled into two classes: *Safe* and *Unsafe*, based on the computed severity threshold.

B. Safety Prediction Model

The core component of the system is the safety prediction model, which uses the XGBoost (Extreme Gradient Boosting) algorithm. XGBoost is selected due to its efficiency, scalability, and strong performance on structured datasets.

The model is trained on historical crime data with multiple features representing different crime categories. It learns pat- terns and relationships between crime frequency and regional safety levels. During prediction, the model classifies a given region as either *Safe* or *Unsafe*.

The dataset is divided into training and testing sets to evaluate model performance. Standard evaluation metrics such as accuracy are used to assess the effectiveness of the model.

C. Crime Hotspot Visualization

To provide spatial awareness of high-risk areas, the system incorporates an intensity-weighted heatmap rendered using the Leaflet.js mapping library and its Leaflet.heat plugin. Each district is represented by its geographical coordinates (latitude and longitude), with the total crime count serving as the spatial weight or intensity value.

The heatmap algorithm aggregates nearby high-intensity data points and renders them using a color-coded gradient overlay: red regions indicate high crime density (hotspots), yellow represents moderate risk zones, and green indicates comparatively safer areas. This approach provides an intuitive, real-time visual representation of crime-prone regions without relying on ML-based clustering algorithms. Additionally, pro- portional circle markers are overlaid on the map, where the circle radius scales with the total crime volume of the district, reinforcing hotspot identification visually.

This visualization aids women tourists in identifying unsafe zones at a glance and planning safer travel routes accordingly.

D. Crime Reporting Module

The system includes a crime reporting interface that allows users to submit incidents either with their identity or anony- mously. This feature encourages wider participation while maintaining user privacy.

Submitted reports are first processed by a pre-trained AI model that detects and filters fake or misleading submissions, ensuring the integrity of crowd-sourced data before it influ- ences safety predictions or officer assignments. Reports are then reviewed by the system administrator and assigned to relevant authorities or safety officers. This ensures a structured and accountable handling of complaints.

E. SOS Emergency Feature

To enhance real-time safety, *SafeRoute* incorporates an SOS emergency feature. When activated, the system sends an immediate alert along with the user's location details to predefined emergency contacts or concerned authorities. This feature is designed to provide quick assistance during critical situations, complementing the predictive capabilities of the system with real-time response support.

F. Chatbot Assistance

A Chatbot model is integrated into the system to assist users in navigation and interaction. The chatbot can respond to common queries, guide users on how to report incidents, and provide general safety-related information.

G. System Workflow

The overall workflow of the system can be summarized as follows:

1. Historical crime data is collected and preprocessed
2. The safety prediction model is trained and used to classify regions
3. Crime hotspots are visualized on an interactive heatmap using district-level crime intensity weights

4. Users interact with the system through the reporting and chatbot modules
5. Submitted reports are validated by a pre-trained AI model for fake report detection before further processing
6. Reports are reviewed by the administrator and forwarded for further action
7. The SOS feature provides immediate assistance during emergencies

The methodology combines data-driven prediction with user-centric features to create a practical safety system. By integrating machine learning models with reporting and emergency functionalities, *SafeRoute* provides both preventive insights and real-time support, making it a comprehensive approach toward improving public safety for women tourists.

IV. System Architecture and Implementation

A. Architecture

The system follows a four-layer architecture:

1. **Input Layer:** Citizens, tourists, officers, and admins interact through role-specific web portals. Anonymous users may submit reports without registration.
2. **Application Layer:** Django handles routing, authenti- cation, form processing, and API endpoints. Role-based access control (RBAC) enforces data visibility rules per user type.
3. **ML Processing Layer:** Trained scikit-learn and XG- Boost models are loaded at runtime for inference. A pre-trained AI model performs fake report detection on submitted incident reports. Crime hotspot visualization is handled by the Leaflet.js library and its Leaflet.heat plugin, which renders intensity-weighted heatmaps using district-level crime counts as spatial weights.
4. **Data Layer:** SQLite (development) / MySQL (produc- tion) stores user accounts, crime reports, officer profiles, and model outputs.

B. Role-Based Portals

- **Citizen / Tourist Portal:** File reports, view safety scores, access the AI chatbot, submit anonymous reports, view the crime hotspot heatmap for identifying high-risk zones and planning safer routes. Tourists can additionally trig- ger the SOS emergency feature to send an immediate alert with live location details to assigned safety officers or emergency contacts.
- **Police Officer Portal:** View assigned cases, update case statuses, view personal performance metrics and charts.

- **Admin Portal:** Approve officer and citizen registrations (with liveness verification review), manage all crime

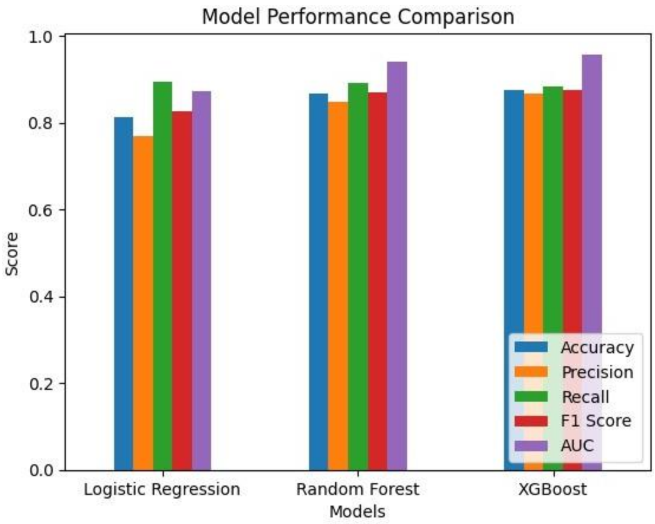

Fig. 1. Model Performance Graph.

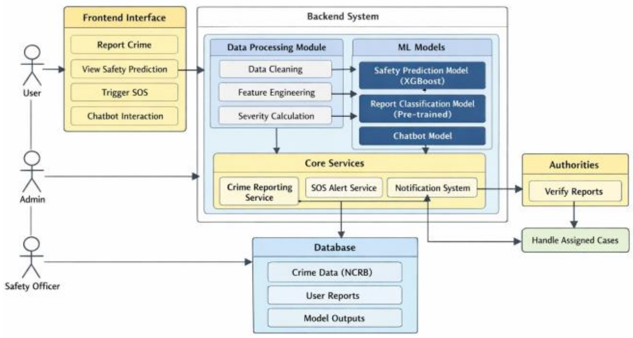

Fig. 2. Safe Route System Architecture Diagram.

Reports, review flagged fake reports identified by the pre- trained AI model, access system-wide analytics.

V. Results and Discussion

The proposed system, *SafeRoute*, was rigorously evaluated based on the predictive performance of its underlying machine learning models and the practical efficacy of its user-facing modules. The experimental results demonstrate the system's capability to extract actionable insights regarding regional safety levels while seamlessly facilitating user-driven reporting and emergency assistance.

A. Safety Prediction Results

The safety prediction engine, driven by the XGBoost algo- rithm, was trained and validated on National Crime Records Bureau (NCRB) data spanning 2019 to 2022. The model was tasked with establishing complex, non-linear correlations between distinct crime categories and regional safety indices. Upon evaluation against a hold-out test set, the XGBoost model achieved a robust accuracy of 87.46% and an impressive Area Under the Curve (AUC) of 0.9572. Furthermore, it recorded a precision of 0.8677, a recall of 0.8837, and an F1-Score of 0.8756, highlighting its superior capability in dis- tinguishing between *Safe* and *Unsafe* regional classifications compared to traditional baseline models.

Analysis indicates that the frequency of severe offenses acts as a primary determinant in the prediction outcomes, yielding consistent results across various districts. However, it must be acknowledged that the model's efficacy is inherently tied to the completeness of the underlying dataset. Given the real- world challenges of underreported or inconsistent crime data, these predictions are designed to serve as highly indicative probabilistic assessments rather than absolute certainties.

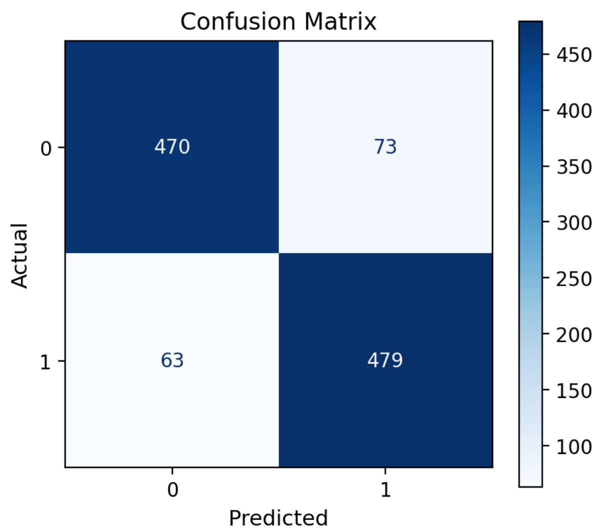

Fig. 3. Confusion Matrix – XGBoost Safety Prediction Model.

TABLE I MODEL PERFORMANCE METRICS.

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**AUC**
Logistic Regression	0.8129	0.7686	0.8948	0.8269	0.8716
Random Forest	0.8654	0.8473	0.8911	0.8687	0.9399
XGBoost	**0.8746**	**0.8677**	**0.8837**	**0.8756**	**0.9572**

B. Crime Report Classification Results

To evaluate the text-processing pipeline, the pre-trained classification model was tested on its ability to accurately categorize unstructured, user-submitted incident reports. The model exhibited reliable performance in extracting contextual features to determine both the nature and severity of the reported incidents from raw textual input.

This automated classification drastically reduces the ad- ministrative overhead required to triage large volumes of incoming reports, empowering law enforcement administrators to prioritize high-risk cases efficiently. While the model per- forms exceptionally well with well-articulated input, a noted limitation involves a degradation in accuracy when processing vague, highly colloquial, or incomplete reports a standard constraint in text-based classification architectures.

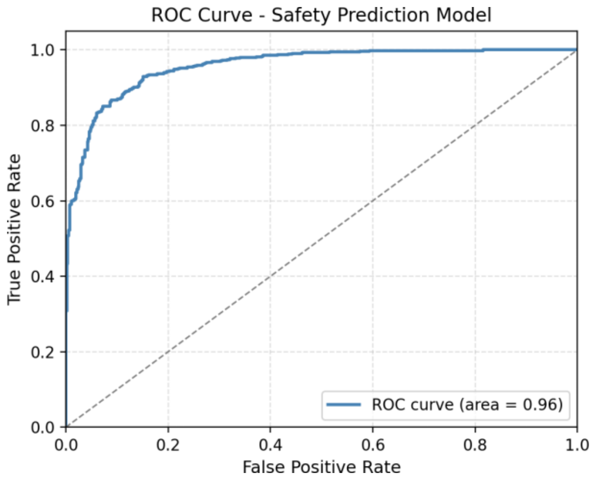

Fig. 4. ROC Curve – Safety Prediction Model (AUC = 0.96)

TABLE II CLASSIFICATION REPORT.

Class/Metric	Precision	Recall	F1-Score	Support
0	0.88	0.87	0.87	543
1	0.87	0.88	0.88	542
Accuracy			*0.87*	*1085*
Macro Avg	0.87	0.87	0.87	1085
Weighted Avg	0.87	0.87	0.87	1085

C. System Functionality Evaluation

Beyond algorithmic performance, the holistic system was subjected to usability and functional testing across its primary modules:

- **Crime Reporting Module:** Facilitates the frictionless submission of complaints. The inclusion of an anony- mous reporting protocol successfully lowers the barrier to entry, thereby encouraging broader citizen and tourist participation.

- **Crime Hotspot Heatmap:** Provides an intuitive, real- time spatial overview of high-risk districts using intensity-weighted color gradients, enabling tourists to make informed route decisions before and during travel.
- **SOS Emergency Feature:** Provides instantaneous alert broadcasting, enabling rapid response deployment during critical, time-sensitive situations.
- **AI Chatbot Interface:** Delivers foundational user as- sistance and triage, significantly improving the human- computer interaction (HCI) experience by guiding users intuitively through the platform's features.

These integrated features collectively elevate the accessibil- ity and practicality of the *SafeRoute* ecosystem, validating its readiness for real-world deployment.

D. DISCUSSION

The comprehensive results indicate that *SafeRoute* effec- tively bridges the gap between predictive analytics and user- centric public safety tools. The XGBoost-powered prediction model serves as the analytical core, providing continuous, data-driven insights into regional vulnerabilities. The heatmap- based hotspot visualization translates district-level crime data into an immediately actionable spatial interface for tourists, complementing the prediction engine with geographic context. Concurrently, the robust report classification pipeline and interactive citizen modules transform these insights into a highly functional, responsive framework. Ultimately, *SafeR- oute* proves to be a viable, scalable solution for modernizing crime reporting and proactive safety monitoring for women tourists.

VI. CHALLENGES AND LIMITATIONS

A. Challenges

- **Data quality:** NCRB data, while authoritative, contains reporting inconsistencies across districts due to varying local recording practices.
- **Class imbalance:** Severe crime categories are under- represented in available datasets, requiring careful class weighting during model training.
- **Geocoding accuracy:** User-submitted location data de- pends on browser geolocation precision, which may vary across devices and network conditions.
- **Fake report labeling:** Ground-truth labels for fraudulent reports were partially generated through heuristic rules, which may introduce labeling noise.

B. LIMITATIONS

- *SafeRoute* is a decision-support system and does not function as a real-time emergency dispatch platform.
- Model performance is dependent on the quality and volume of training data; smaller districts with fewer reports may yield less reliable predictions.
- The hotspot visualization is based on historical aggregate crime data and does not reflect live or real-time incident streams.
- Deep identity verification for user accounts is not imple- mented in the current version.
- No integration with government emergency dispatch APIs or telecom services is available in the current release.

VII. CONCLUSION

This paper presented *SafeRoute*, a machine learning-based system designed to predict regional safety levels and support crime reporting for women tourists using NCRB data. The core component, based on the XGBoost algorithm, effectively classifies regions as *Safe* or *Unsafe* by analyzing multiple crime-related features. In addition to prediction, the system integrates a pre-trained AI model to validate and filter fake or misleading user-submitted reports, and a heatmap-based hotspot visualization using Leaflet.js and Leaflet.heat that renders district-level crime intensity as spatially meaningful color overlays enabling tourists to identify high-risk zones and plan safer routes. The crime reporting module supports anonymous reporting to encourage public participation.

The inclusion of an SOS emergency feature enables real- time alert generation during critical situations, while a basic chatbot enhances user interaction and accessibility. Although the system provides valuable insights, its performance is dependent on the quality of available data and should be interpreted as indicative. Overall, *SafeRoute* demonstrates a practical and scalable approach to combining machine learning with user-driven features to enhance public safety for women tourists.

REFERENCES

1. National Crime Records Bureau, “Crime in India 2021,” Ministry of Home Affairs, Government of India, New Delhi, India, 2022. [Online]. Available: https://ncrb.gov.in
2. S. G. Shalini et al., “Crime Prediction Using Random Forest Model,” *IJETT*, vol. 72, no. 3, pp. 55–61, 2025.
3. B. Patel and M. C. Zala, “Crime Against Women Analysis & Prediction in India Using

Supervised Regression," *IRJET*, vol. 9, no. 7, pp. 1224– 1231, 2022.

4. M. Mwaniki et al., "Crime Prediction Using Decision Trees and Random Forest: A Comparative Study," *IJCSIT*, vol. 11, no. 2, pp. 150–158, 2023.
5. B. Subha K. et al., "AI-Based Integrated Women Safety System: Real- Time Monitoring and Emergency Response," *IEEE Access*, vol. 13, pp. 121560–121575, 2025.
6. S. A. P. N. Kavala et al., "Crime Analysis and Prediction Using Neural Networks," *8th Int. Conf. on Advanced Computing and Communication Systems (ICACCS)*, IEEE, 2022.
7. A. Srivastava and P. Singh, "Spam Detection Using Natural Language Processing," *Journal of Applied Science and Education (JASE)*, vol. 4, issue 2, 2024.
8. S. Surana, J. Chekkala, and P. Bihani, "Chatbot based Crime Registration and Crime Awareness System using a Custom Named Entity Recognition Model," *IRJET*, vol. 8, issue 4, 2021.
9. K. Gautam, S. Bhujbal, V. Marathe, and N. Naik, "FIR Registration Using Chatbot," *IRJET*, vol. 7, issue 5, 2020.
10. S. Sharma and S. C. Dronavalli, "Data Analysis and Visualization of Crime Data," *IS&T Int. Symp. on Electronic Imaging*, 2024.
11. A. Jain et al., "A Smart System for Fake News Detection Using Machine Learning," *IEEE Int. Conf. on Information, Communication and Computing Technology (ICICCT)*, 2019.