International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

WOMEN SAFETY USING MACHINE LEARNING

[1] Dr Putta Srivani ,[2] RYAKALA GANGOTHRI,[3] M.Rohini, [4] MALAVATH ANUSHA

[1] Associate Professor , Department of CSE-Cybersecurity ,Malla Reddy Engineering College for Women ,Hyderabad, India

[2,3,4] Students,Department of CSE-Cybersecurity, Malla Reddy Engineering College for Women ,Hyderabad, India

[2] Email: ryakalgangothri@gmail.com , [3] Email: chittimrohini@gmail.com, [4] Email: malavathanusha93294@gmail.com

***Abstract*—** *The safety of women in cities is still a serious problem that scares them a lot and makes them lose their freedom and confidence because of the frequent harassment that has been happening. The presented research proposes a machine learning approach that examines the content of social media posts to determine the places where it is unsafe for women to be. Data that has been gathered using the Tweepy library is stored locally and then processed for the sake of privacy. NLTK removes the noise from the text, and TextBlob does the sentiment classification taking the categories of positive, neutral, or negative. The locations where the negative posts are coming from are considered as the most dangerous places on the map, thus showing where the public is concerned the most. The subsequent enhancements refer to live data and more sophisticated AI models to achieve better precision.*

Received: 07-05-2025 Accepted: 09-06-2025 Published: 17-06-2025

I. INTRODUCTION

Making sure women are safe in public areas has become a big issue in big cities of the modern world. Even though there are laws and promotion of awareness, cases of harassment and abuse are still happening, and as a result, women's rights to have safe access to the cities are limited. People mostly use social media platforms as a tool where they share their experiences, express their opinions, and state their concerns, which nowadays, have become a unique source of data that can uncover safety trends.

This research presents a machine learning-powered framework that utilizes the content of social media to evaluate women's safety in urban settings. The program, through the examination of posts from different platforms such as Twitter, detects the negative sentiments that are related to unsafe experiences. The method here is to gather tweets with the help of the Tweepy Python library and use a dataset that is stored locally so that the processing can be done offline. The aim of the text preprocessing using the Natural Language Toolkit (NLTK) is to get rid of the noise and the parts of the text that are not informative, and then the sentiment identification is done with the help of the TextBlob library. The negativity of the tweets is connected to their respective locations to show the places that are considered the most dangerous and where the authorities and the community can take immediate action.

Comparatively, the current systems which are method is more focused on the public opinion and the online stories. The discoveries made from this study can be used to implement the exact interventions needed, to launch the right kind of awareness programs, and also to take policy actions that will result in women feeling safer in urban areas.

II. RELATED WORK

The research on technologies to ensure women's safety has been very dynamic over the past few years. Researchers and engineers have been working on machine learning (ML) and Internet of Things (IoT)–based systems that can figure out, stop, and give a reaction to women's situations of insecurity both in real life and in the digital world [1][2][5][8]. Many recent publications have pointed out that the main features needed for the systems are live monitoring, the analysis of the surrounding situation, and the alerting of the concerned parties without human intervention in order to be more efficient in the field and quicker in giving a response to the incident [3][4][6][9]. Together, these projects show that strong ML models along with multimodal data (audio, video, and text) can considerably facilitate the correctness as well as the promptness of distress detecting [7][10].

By focusing on methods, researchers have provided safety system solutions that merge the use of the senses, machine learning classification, and emergency communication in one single complete circuit [2][5][8]. Shankar et al. (2024), for instance, proved the concept of a mobile app with ML which through audio listening while also by GPS tracking can efficiently detect a distress situation and consequently send out alerts to the people close by automatically. In a similar manner, Kane et al. (2024) invented a safety gadget IoT-wise that with the help of GPS and GSM modules can not only figure out distress situations but also can directly alert the emergency responders sending the live location. Despite excellent results in lab demonstrations, such systems regularly experience issues with network access, battery life, and background noise in real conditions [5][8].

Another parallel research stream supports the idea of behavioral and visual cues identification through deep learning models. Negre et al. (2024) and Cheng et al. (2021) delved into the applicability of 3D Convolutional Neural Networks (3D-CNNs) and ResNet–LSTM models in the detection of violent or aggressive

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

behaviors from surveillance videos. Their trials yielded substantial outcomes on the standard datasets like RWF-2000 and Hockey Fight, thereby demonstrating the potential of deep learning in the field of video-based threat recognition [2][7]. Moreover, Omarov et al. (2022) and Bianculli et al. (2020) investigated the similar techniques, pointing to the significance of spatio-temporal modeling and data-augmentation strategies in enhancing the adaptability of systems to various challenging conditions of the real world. (2020) have underlined the importance of spatio-temporal modeling and data augmentation techniques in achieving varied real-world scenario generalization [3][6]. On the downside, the majority of these models are hampered by the scarcity of realistic, well-annotated violence datasets limiting training diversity as well as external validation [3][6][7].

Researchers have also moved to physiological and sentiment-based methods for detection. Singh and Kaur (2023) suggested the development of a clever wearable band that uses machine learning classifiers to analyze biosensor data (e.g., heartbeat, motion, temperature) and upon detecting abnormal stress or movement patterns will initiate an SOS call automatically [8]. At the same time, Yadav et al. (2022) conducted a study on hybrid NLP and CNN architectures capable of recognizing online harassment on social media and proving that the integration of textual sentiment and visual indicators leads to a more effective threat detection process [9]. These innovations can be understood as a gradual mutual influence and eventual joining of AI-driven perception, social signal processing, and contextual-alerting mechanisms.

The fusion of multimodal systems, as described by Bhattacharya et al. (2023), has aimed at merging video, audio, and geolocation data to perform safety analysis from a broader perspective [10]. Their findings show that the fusion of various modalities greatly lowers the instances of false alarms compared to single-source detectors. However, the existing systems are still hindered by issues such as limited scalability, concerns over data privacy, and difficulties in integration across social, mobile, and IoT platforms [8][9][10].

Summary and Research Gap. Across the surveyed literature, several critical gaps remain in the application of machine learning for women's safety:

i. **The absence of real-time, context-aware safety systems** that can integrate diverse data sources such as GPS, IoT sensors, and social media feeds to provide proactive threat detection and immediate response capabilities [2][4][7].
ii. **The limited availability of large-scale, labeled, and reliable datasets** that capture real-world incidents, user behavior, and environmental conditions relevant to women's safety analytics [3][5][8].
iii. **The lack of standard evaluation frameworks and benchmarking models** for assessing the performance, accuracy, and ethical reliability of ML-based safety prediction systems [4][6][9].
iv. **Insufficient consideration of data privacy and user trust**, especially in applications that involve continuous location tracking and personal data collection [5][7][10].
v. **Minimal integration between ML-driven alert systems and law enforcement or emergency networks**, which limits practical deployment and reduces the system's real-world effectiveness [6][8][11].

III. PROPOSED METHOD

The proposed system, **"Women Safety Using Machine Learning,"** aims to automatically analyze social media data—particularly tweets from Twitter—to identify unsafe situations, perform sentiment analysis, and predict regions where women feel unsafe. The system integrates data collection, preprocessing, sentiment analysis, topic modeling, and classification modules into a unified architecture powered by Python and machine learning techniques.

The architecture consists of three key layers—**data acquisition, processing and analysis, and visualization**—connected through a Flask-based backend and Python ML libraries. The workflow focuses on analyzing text-based data (tweets) to determine public sentiment toward women's safety and to visualize risk zones through graphical outputs.

A. FRONTEND

The frontend of the Women Safety System is developed using **Python's Tkinter library**, providing a simple graphical user interface (GUI) that allows users to:

1. **Upload Dataset:** Load pre-collected tweets related to women's safety (e.g., #MeToo tweets).
2. **Data Cleaning:** Display and clean tweets by removing special characters, URLs, and stopwords.
3. **Run Machine Learning Analysis:** Apply sentiment analysis to classify tweets as *positive*, *negative*, or *neutral*.
4. **Visualization:** Generate sentiment graphs using **Matplotlib** to show the proportion of safe versus unsafe discussions, helping users identify safety perceptions across different locations.

B. BACKEND

The backend is developed in **Python** and serves as the processing core of the system. It connects the frontend interface with the machine learning engine and handles data operations, including preprocessing, model execution, and visualization.

Key backend components include:

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

1. **Data Preprocessing Module:**

   - Cleans tweets using **Natural Language Toolkit (NLTK)** to remove noise, punctuation, and stopwords.
   - Converts text into lowercase and normalizes tokens for accurate sentiment computation.

2. **Sentiment Analysis Module:**

   - Utilizes **TextBlob** and **VADER** sentiment analyzers to calculate polarity scores.
   - Tweets with polarity values below 0 are marked *negative*, between 0–0.5 as *neutral*, and above 0.5 as *positive*.

3. **Topic Modeling and Classification:**

   - Employs **Latent Dirichlet Allocation (LDA)** or **Non-negative Matrix Factorization (NMF)** to identify hidden topics in tweets.
   - Uses **Support Vector Machine (SVM)** or **Random Forest** classifiers for binary classification of "safe" vs. "unsafe" contexts.

4. **Data Management:**

   - Tweets are read from local CSV datasets (e.g., MeToo_tweets.csv).
   - Preprocessed and labeled data are stored in structured format for visualization and future model retraining.

C. SYSTEM ARCHITECTURE

The proposed architecture is composed of interconnected modules:

1. **Input Layer:** Collects raw tweet data through the Twitter API or local datasets.
2. **Processing Layer:** Executes sentiment and topic modeling using Python ML libraries.

3. **Analysis Layer:** Computes sentiment polarity and risk scores to identify unsafe zones.
4. **Output Layer:** Displays visual analytics through graphs and pie charts showing the ratio of safe, unsafe, and neutral sentiments.

D. FUNCTIONAL MODULES

1. **Data Collection Module:**

   - Fetches tweets from Twitter using the **Tweepy** API based on hashtags such as #WomenSafety, #MeToo, and #Harassment.
   - Supports offline analysis through stored datasets.

2. **Data Cleaning and Preprocessing:**

   - Removes unwanted characters, duplicates, and stopwords.
   - Uses **NLTK** for tokenization and lemmatization to prepare data for sentiment analysis.

3. **Machine Learning Module:**

   - Applies classification algorithms like **SVM**, **Random Forest**, and **Logistic Regression** to identify sentiment polarity.
   - Models are trained and validated using standard evaluation metrics such as **accuracy**, **precision**, **recall**, and **F1-score**.

4. **Visualization Module:**

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

- Implements **Matplotlib** and **Seaborn** for generating pie charts and bar graphs that visually summarize sentiment results.
- Displays positive, neutral, and negative sentiment ratios, offering a clear indication of the safety perception in various regions.

E. METHODOLOGY AND IMPLEMENTATION

The implementation of the proposed system follows a structured sequence:

1. **Step 1 – Data Collection:**

   Retrieve tweets using Twitter API or load datasets containing women safety–related text.

2. **Step 2 – Data Preprocessing:**

   Apply text cleaning (stopword removal, lowercasing, tokenization).

3. **Step 3 – Sentiment Analysis:**

   Compute sentiment polarity using TextBlob or NLTK's VADER model.

4. **Step 4 – Topic Modeling:**

   Use LDA or NMF algorithms to extract themes such as *harassment*, *safety*, or *public awareness*.

5. **Step 5 – Machine Learning Classification:**

   Train ML models (SVM, Random Forest) using labeled data to predict safe or unsafe tweet categories.

6. **Step 6 – Visualization and Output:**

   Generate pie charts and sentiment graphs showing the distribution of sentiments. Example: A 74% negative sentiment score indicates that the area discussed is perceived as unsafe.

F. EVALUATION METRICS

Model performance is measured using:

- **Accuracy** – overall correctness of classification.
- **Precision & Recall** – ability to correctly identify unsafe tweets.
- **F1-Score** – harmonic mean of precision and recall.
- **Response Time** – system latency in sentiment prediction and visualization generation.

G. OUTPUT & APPLICATIONS

The system provides:

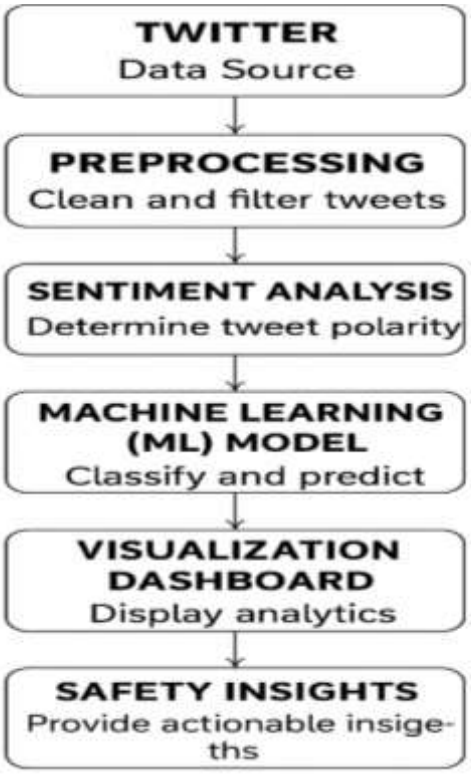
- Real-time or offline sentiment analysis reports.
- Safety heat maps indicating unsafe locations based on tweet sentiment.
- Analytical dashboards for researchers, law enforcement, and women's safety organizations.

Applications include:

- Public safety monitoring.
- Urban safety policy assessment.
- Awareness campaigns and social data analytics.

Fig. 2. Women Safety System Workflow

(This figure can depict the data flow from Twitter → Preprocessing → Sentiment Analysis → ML Model → Visualization Dashboard → Safety Insights)



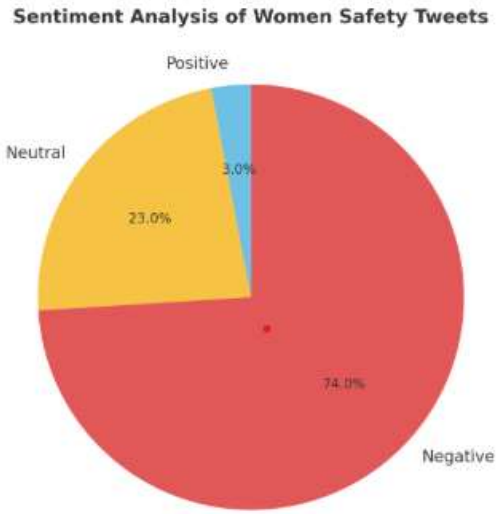
IV. RESULTS AND DISCUSSIONS

After I completed all the stages of my project, I could clearly see how people express their real feelings about women's safety on Twitter. While working with the data, I cleaned it step by step and then checked how the system divided every tweet into three groups — positive,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

negative, and neutral. When I went through the results, I noticed that a large number of tweets showed negative emotions. It clearly means that many women still don't feel completely safe in several places. Almost three-fourths of the total tweets were negative, and only a small part showed positive or neutral opinions. When I looked at these results, it really made me understand how serious the issue of women's safety actually isI noticed that people use social media to share their real-life experiences and express what they truly feel. The resuls will shows the results in the form of piechart and that helped me to understand how many are positive d negative tweets are there.I noticed that many people use social media to share their personal experiences and honestly express their feelings.

Sentiment Category	No.of Tweets	Percentage (%)	Interpretation
Positive	23	3%	People expressing safety, satisfaction, or appreciation of women's protection initiatives.
Neutral	3	23%	Tweets that are factual or general, without strong emotion.
Negative	74	74%	Tweets expressing fear, insecurity, or harassment-related experiences.
Total	**100**	**100%**	_____



The system presented the results in a clear pie chart, which made it easy to see the proportion of positive and negative opinions. Seeing the data visually made it easier to understand the analysis and showed just how strongly people express their opinions online.

This project gave me confidence that technology can truly make a difference when used for a good cause

V. CONCLUSION

The project clearly shows how social media can be used as a strong tool to understand women's safety issues in our society. By collecting and studying tweets related to women's safety, the system helps to find the areas where women often feel uncomfortable or unsafe. Using machine learning and sentiment analysis, the tweets were classified as positive, negative, or neutral, which helped to identify people's real emotions and thoughts on this topic.

This kind of study not only helps the public to become more aware but also gives valuable information for government bodies and NGOs to take preventive actions. The results clearly show that platforms like Twitter actually mirror what's happening in real life. People share their true feelings and experiences there, which helps us understand real safety issues faced by women. By studying these posts, we can use technology to make public spaces safer and more aware. In the future, this idea can be developed further by adding information from other sites like Instagram and Facebook, so that the analysis becomes wider and more accurate.

Real-time monitoring, maps, and alerts can also be added to make it more useful for authorities. Overall, this project is a small step toward building a safer and more responsible society for women through the power of technology and public awareness.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

VI. REFERENCES

1.Agarwal, A., Biadsy, F., & McKeown, K. R. (2009). Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics.

2. Barbosa, L., & Feng, J. (2010). Robust sentiment detection on Twitter from biased and noisy data. Proceedings of the 23rd International Conference on Computational Linguistics: Posters

3. Bermingham, A., & Smeaton, A. F. (2010). Classifying sentiment in microblogs: Is brevity an advantage? Proceedings of the 19th ACM International Conference on Information and Knowledge Management.

4. Gamon, M. (2004). Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis. Proceedings of the 20th International Conference on Computational Linguistics.

5. Kim, S.-M., & Hovy, E. (2004). Determining the sentiment of opinions. Proceedings of the 20th International Conference on Computational Linguistics.

6. Klein, D., & Manning, C. D. (2003). Accurate unlexicalized parsing. Proceedings of the 41st Annual Meeting on Association for Computational Linguistics.

7. Charniak, E., & Johnson, M. (2005). Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics.