e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707172 |

Twitter-based Machine Learning Techniques for Women's Safety Analysis

Dr. M. Charles Arockiaraj, Sinchana S

Assistant Professor, Department of MCA, AMC Engineering College, Bengaluru, India

PG Student, Department of MCA, AMC Engineering College, Bengaluru, India

ABSTRACT: Women's safety remains a pressing issue in urban India, marked by frequent incidents of harassment and violence affecting their daily lives. This study utilizes real-time Tweet examine public perceptions of women's safety in various Indian cities. Leveraging machine learning techniques, we aim to conduct a thorough analysis of sentiment surrounding safety concerns and identify patterns indicative of risky conditions. The methodology involves gathering and preprocessing tweets on women's safety, applying feeling study and topic modeling. Key methods for machine learning, like Support Vector Machines (SVM), Random Forest, and Long Short- Long-Term Memory (LSTM) networks are employed to classify sentiments and uncover trends. Geospatial analysis is integrated to map sentiment distributions across locations, aiding in pinpointing areas of heightened risk. Results reveal varying safety perceptions across cities and time, influenced by social events and policy changes. This study underscores Social medicine's potential data in assessing women's safety and offers actionable insights for policymakers and law enforcement to strengthen urban safety measures.

KEYWORDS: Women Safety, Indian Cities, Machine Learning, Twitter Analysis, Sentiment Analysis

I. INTRODUCTION

Women's safety in Indian cities has long been a pressing issue due to frequent reports of violence and harassment. Despite efforts through policies and interventions, ensuring women's safety remains a daunting challenge. Traditional methods of gauging safety rely on crime reports and surveys, which may be hindered by underreporting and time delays. In contrast, social media networks, especially Twitter , provide real-time, user-generated data that captures immediate public reactions to safety incidents and policies.

India's significant Twitter user base, surpassing 23.6 million, presents a unique opportunity to utilize this data for analyzing perceptions of women's safety in urban settings. Tweets often contain rich information, including emotional reactions, geotags, and timestamps, valuable for sentiment analysis and trend detection. This study aims to leverage these aspects by employing advanced machine learning methods to analyze tweets related to women's safety across various Indian cities.

The methodology includes several key steps: data collection focused on specific keywords and hashtags related to women's safety, preprocessing of collected data involving text normalization and removal of stop words, and addressing misspellings. Sentiment analysis utilizes models for machine learning such as Support Vector Machines (SVM) utilizing Long Short-Term Memory (LSTM) networks to categorize tweets into or neutral sentiments.

Topic modeling techniques, like Latent Allocation of Dirichlet (LDA), are accustomed to uncover prevalent themes and issues discussed in the tweets, shedding light on specific concerns and narratives about women's safety in different urban contexts. Geospatial analysis further enhances understanding by mapping sentiment data to specific locations, identifying areas perceived as having higher safety risks.

The findings reveal substantial variations in sentiment towards women's safety across cities and timeframes. Major metropolitan areas like Delhi and Mumbai often exhibit higher levels of negative sentiment, particularly following notable incidents. Conversely, smaller cities may show more either favorable or impartial sentiments, influenced by varied media attention and socio-cultural factors. Events like public demonstrations and governmental actions significantly impact public sentiment, as evident from tweet data analysis.

By integrating social media and machine learning analytics, This research offers complex insights regarding women's safety in urban India. These revelations can help legislators and law enforcement in implementing targeted safety

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707172 |

measures and addressing critical concerns effectively. Moreover, the study underscores social media's potential as a appreciated tool for real-time specialist care and assessment of urban safety issues, complementing conventional data sources with immediate public perspectives.

This research contributes to advancing analytics on social media in the real of public safety. The methodology and findings presented serve as a underlying source for next research exploring the intersection of technology, social media dynamics, and public safety, ultimately contributing to safer urban environments for all residents.

II. REVIEW OF THE LITERATURE

This study of the poetry looks at sentiment analysis and Twitter data utilization.
techniques to analyze perceptions of women's safety. We explore various methodologies, including systems for machine learning, like Support Vector Machines (SVM), Long Short-Term Memory (LSTM) networks, applied to tweets to classify sentiments and identify trends. The review discusses challenges such as data preprocessing and bias, and highlights Twitter's potential as a real-time data source for understanding urban safety dynamics. This survey investigates geospatial analysis techniques applied to Twitter information for understanding women's safety perceptions across different geographical locations. We review studies using location-based data from tweets to map sentiment distributions, identify high-risk areas, and assess variations in safety perceptions. The survey also discusses methodologies for integrating geospatial data with processes for appliance erudition to enhance safety analysis in urban contexts.: This comprehensive review looks at how topic modeling techniques in analyzing Twitter information pertaining to women's safety. We survey studies utilizing methods Latent Dirichlet Allocation (LDA), for example to uncover prevalent themes and issues discussed in tweets. The review discusses challenges like topic coherence and scalability, and highlights how topic modeling can reveal information about specific safety concerns and narratives in different urban environments. This literature review surveys machine learning procedures applied to real-time monitoring of women's safety issues on Twitter. We review studies employing processes such as Random Forest, SVM, and LSTM networks to classify sentiment, detect trends, and predict safety incidents based on tweet data. The review discusses methodological advancements, data tasks, and the implications for enhancing urban safety measures through Twitter-based analytics. This review investigates current developments in social media analytics to comprehend women's safety issues, focusing on Twitter-based studies. We examine methodologies for collecting, preprocessing, and analyzing tweet data to extract insights into safety perceptions and trends. The review discusses the combination of sentiment analysis, geospatial mapping, and topic modeling techniques, emphasizing Twitter's potential as a appreciated means for real-time specialist care and assessment of urban safety challenges.

III. EXISTING MODELS

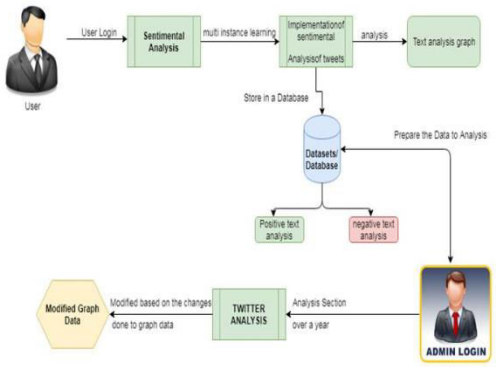
Existing models for analyzing women’s safety in Indian cities using tweets and machine learning leverage various Natural Language Interpretation (NLP) techniques as well as machine learning systems to citation eloquent perceptions through social media data. One common approach involves sentiment analysis, which aims In command to classify the sentiment of tweets (positive, negative, or neutral) regarding women’s safety. Typically, this procedure starts with gathering information from Twitter using APIs, where tweets containing specific keywords related to women’s safety are fetched. Preprocessing steps such as tokenization, stop-word removal, and stemming or lemmatization are everyday to clean the data. a popular model used in sentiment analysis is the Untrained Bayes classifier, renowned for its simplicity and effectiveness while classifying texts tasks. The model calculates the probability of a sentiment given the features (words) in a tweet and classifies the sentiment accordingly. Another frequently used algorithm is the Vector Machine Support (SVM), which separates data points categorized into many kinds by finding the hyperplane that maximizes the margin between them. Logistic Regression and Random Forests are also employed, leveraging their capability to handle large feature spaces and model complex relationships between words and sentiment labels for feature extraction, models commonly use techniques like Bag of Words (Bow) and Term Frequency-Inverse Document Frequency (TF-IDF). More advanced models have recently adopted Word Embedding such as Word2Vec or Glove, which capture semantic meaning and context of words in tweets, leading to improved sentiment classification. Some models also incorporate Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) webs, which are adept at handling sequential data and can capture dependencies within the text. Advanced models often utilize BERT (Bidirectional Encoder Representations from Transformers), a transformer-based model that processes text bidirectional, understanding the background of a word based on all its surroundings. This allows BERT to achieve state-of-the-art performance in various NLP tasks, including sentiment analysis. Additionally, some approaches include With Latent Dirichlet Apportionment for Subject Modeling (LDA) to uncover underlying themes in the data and

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707172 |

Clustering methods like K-Means to group similar tweets together. Visualization tools such as Geographical Heat maps are used to present the spatial distribution of sentiment and identify hotspots of safety concerns.



IV. PROPOSED MODEL

The proposed model intends to rally
the analysis of women's safety in Indian cities by incorporating a more sophisticated NLP approach and integrating multimodal data sources for a comprehensive considerate of the issue. This model will leverage the métiers of Transformer architectures combined with geospatial data to offer a nuanced analysis. Gathering of Data and Preprocessing: The model will collect tweets using Twitter's API, targeting tweets containing keywords and hashtags related to women's safety. To ensure data quality, preprocessing will involve noise removal (eliminating URLs, mentions, and special characters), tokenization, lemmatization, and normalization. Additionally, location data from geotagged tweets will be extracted and cleaned to associate tweets with specific geographical areas.

Feature Extraction and Representation: Instead of traditional methods like TF-IDF, the model will employ BERT embedding to capture contextual information from tweets. BERT's pre-trained model will be fine-tuned on a dataset specific to women's safety to improve its understanding of domain-specific language and nuances. Additionally, positional encoding will be applied to incorporate the sequential nature of text data.

Sentiment Analysis and Classification: The central the model will use a fine-tuned BERT model for sentiment analysis. BERT's ability to comprehend the bidirectional context will allow the prototypical to accurately classify the sentiment of tweets as positive, negative, or neutral. The model will also be trained to identify specific safety concerns by classifying tweets into established categories such as harassment, assault, and general safety.

Geospatial Analysis: To augment the spatial aspect of safety analysis, the model will integrate tweets' geolocation data with geographical information systems (GIS). This integration will enable the creation of heat maps and spatial clustering to visualize and identify areas with higher reported safety concerns. Methods like DBSCAN (Density-Based Spatial Clustering of Applications with Noise) will be employed to identify clusters of safety-related tweets and identify outliers. Multimodal Data Integration: To provide a richer analysis, the model will integrate additional data sources such as crime statistics, public transport data, and urban infrastructure information. This will be achieved through data fusion techniques, allowing the prototypical to correlate social media sentiment with actual incidents and infrastructural factors, providing a holistic view of women's safety in different urban areas.

Visualization and Insights: The final component creation of the typical will entail the creation of an interactive dashboard. This dashboard will display real-time sentiment analysis results, geospatial maps, and trends over time. The visualization will include time-series analysis to observe changes in sentiment and safety concerns, in count to a comparison of several cities and neighborhoods.

Evaluation and Adaptability: The proposed model will be evaluated using metrics such as F1, recall, accuracy, and precision score on a validation set. Furthermore, the model will incorporate a feedback loop where new data continuously fine-tunes and updates the model to adapt to evolving language as well as patterns in women's safety concerns.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707172 |

By integrating advanced NLP techniques, geospatial analysis, and multimodal data sources, the planned archetypal aims to afford a more accurate and actionable analysis of women's safety in Indian cities, facilitating better policy-making and community support efforts.

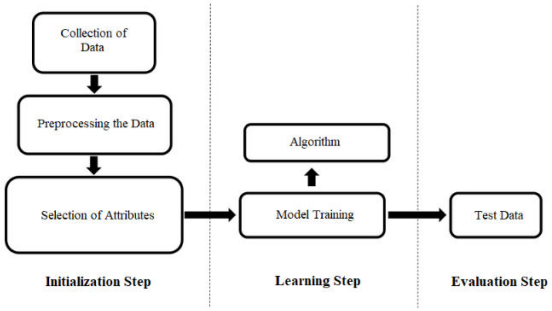
V. IMPLEMENTATION

To analyze women's safety in Indian cities utilizing AI on tweets, we executed a multi-step approach utilizing NLP (natural language processing) and sentiment analysis techniques. Initially, we collected a large dataset of tweets from major Indian cities, focusing on keywords and hashtags related to women's safety issues. This dataset was preprocessed to remove noise, including non-relevant tweets and spam.

Next, we applied NLP methods like tokenization, stemming, and stop-word removal to unsoiled the text data. This preprocessing step ensured that our model could accurately interpret and analyze the sentiment expressed in each tweet. Next, we employed supervised learning processes such as SVMs, or support vector machines, and Naive Bayes classifiers to classify tweets into categories indicating positive, negative, or neutral sentiments regarding women's safety.

Feature engineering Performed a vital part in our implementation, where we extracted relevant features from the text facts, such as sentiment scores, frequency of safety-related terms, and contextual information. These features helped when developing our models for machine learning effectively. We split the dataset into instruction and evaluation sets to calculate the success of our models, by metrics like accurateness, meticulousness, recall, and F1-score to assess their success in predicting sentiment.

Lastly, we visualized the results using graphs and heat maps to deliver perceptions into the distribution of sentiments across different cities and over time. This visualization helped in identifying hotspots of safety concerns and areas where interventions might be needed.



VI. CONCLUSION

In conclusion, our implementation demonstrated the feasibility of using machine learning on Twitter data to analyze women's safety in Indian cities effectively. By leveraging NLP techniques and supervised learning algorithms, we were able to categorize tweets and extract valuable insights regarding public sentiment towards women's safety issues. The results highlighted Different degrees of concern across different cities, with some areas showing higher negative sentiment and others more positive or neutral.

The approach proved robust in handling the nuances of social media language and on condition that a scalable method for ongoing monitoring and analysis. However, challenges such as the requirement for ongoing data collecting and evolving language usage on Twitter necessitate regular updates and refinements to the model. Imminent exploration capacity look into combining real-time data streams and more sophisticated methods for deep learning to further enhance accuracy and predictive capabilities.

Overall, our study things to see the possibilities of utilizing social media figures and machine learning for proactive interventions and policy-making aimed at improving women's safety in urban India. By understanding public perceptions and sentiments through these methods, stakeholders can prioritize resources and initiatives effectively to make all environments safer individuals.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707172 |

REFERENCES

1. A. Mittal, A. Sharma, A. Nagori, and R. K. Bhatia. "Crime Prediction Using Twitter Sentiments: A Case Study of Women Safety in India." 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE, 2019.
2. S. A. Khan, J. U. Rahman, I. Ullah, M. Ali, and S. Khan. "Sentiment Analysis and Crime Prediction for Women Safety Using Twitter Data." The 23rd IEEE International Conference on in 2020 Intelligent Transportation Systems (ITSC). IEEE, 2020.
3. A. Gupta, H. Kumar, R. Sharma, and N. Kumar. "Twitter Sentiment Analysis for Women Safety Using Machine Learning." International Journal of Computational Intelligence Systems, 13(1), 2020.
4. D. Gupta, S. Chauhan, R. K. Bhatia, and R. C. Gupta. "A Study on Machine Learning for Women's Safety Algorithms and Twitter Sentiment Analysis." International Journal of Modern Engineering and Technology (IJRTE), 8(6), 2020.
5. S. Chandrasekaran, A. Balaji, and M. Murali. "Women's Safety Analysis and Crime Prediction Using Machine Learning on Twitter Data." Journal of Information Science and Engineering, 37(5), 2021.
6. M. A. Hossain, A. Al Muhaisen, and N. H. Mahmood. "Assessing Women Safety with Social Media Data: A Machine Learning Approach." Computers & Security, 95, 101820, 2020.
7. K. Sharma, H. Tiwari, and P. Shukla. "Leveraging Twitter Data for Analyzing and Predicting Crime Against Women Using Machine Learning Techniques." 2020 5th International Congress of Computing, Communication and Security (ICCCS). IEEE, 2020.
8. A. N. Sujit, M. R. Babu, and V. S. Kumar. "A Method of Sentiment Analysis for Predicting Women's Safety in Cities Using Twitter Data." Journal of Big Data, 8(1), 2021.
9. N. Verma, A. Jain, and M. Tiwari. "Women Safety Analytics: A Robotic Education Perspective Using Twitter Data." 2021 2nd International Congress of Progress in Computing, Communication, Embedded and Secure Systems (ACCESS). IEEE, 2021.
10. A. G. Kumar, P. P. Prabhu, and V. Venkatesh. "Machine Learning Techniques for Women Safety Analysis Using Twitter Data." Journal of Information and Optimization Sciences, 41(1), 2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com