,

INVESTIGATING WOMEN'S SAFETY IN INDIAN CITIES: A MACHINE LEARNING ANALYSIS OF TWITTER SENTIMENTS

[1]SWETHA ARRA, [2]SUSHRITHA AMBATI, [3]RAJASHEKAR KANDAKATLA, [4]KODITHALA VISHWAS, [5]KANNAVENA RAKESH

[123]Assistant Professor, Department of CSM, Vaagdevi College of Engineering, Warangal, Telangana

[45]Students, Department of CSM, Vaagdevi College of Engineering, Warangal, Telangana

ABSTRACT:

This study presents a comprehensive analysis of women's safety in Indian cities through the lens of social media, utilizing machine learning techniques to interpret and analyze tweets related to safety concerns. With the increasing prevalence of gender-based violence and harassment, understanding public sentiment and the specific challenges faced by women is crucial for policy-making and community support initiatives. By employing natural language processing (NLP) algorithms, we systematically extracted and categorized relevant tweets to identify prevalent themes and sentiments regarding women's safety. Our methodology includes data collection from Twitter, sentiment analysis, and the implementation of machine learning models to classify tweets based on their relevance to safety issues. The findings reveal significant trends and patterns, highlighting areas of concern and public perception of safety in various urban regions. This research not only contributes valuable insights into the discourse surrounding women's safety in India but also underscores the potential of machine learning and social media analytics as powerful tools for social research and policymaking. Ultimately, this study aims to foster informed discussions and drive initiatives focused on enhancing women's safety and security in urban environments.

I.INTRODUCTION:

The safety of women in urban environments has become a pressing social issue, particularly in countries like India, where incidents of gender-based violence and harassment have been on the rise. As cities continue to expand and urbanization accelerates, understanding the complexities surrounding women's safety is crucial for developing effective policies and community interventions. Social media platforms, especially Twitter, have emerged as valuable sources of real-time public sentiment and discourse, reflecting the lived experiences and concerns of individuals. This study leverages Twitter data to analyze the perceptions and discussions related to women's safety in various Indian cities, providing a unique lens to understand public attitudes and societal challenges.

Recent advancements in machine learning and natural language processing (NLP) enable the systematic analysis of large datasets, allowing researchers to extract meaningful insights from unstructured text. By applying these techniques to tweets about women's safety, this research aims to identify prevalent themes, sentiments, and patterns that may not be readily apparent

through traditional qualitative methods. The analysis focuses on categorizing tweets based on their relevance to safety issues, encompassing various dimensions such as harassment, violence, and public safety measures.

The significance of this research extends beyond academic inquiry; it seeks to inform policymakers, social organizations, and communities about the current state of women's safety in urban India. By synthesizing social media narratives with machine learning methodologies, this study aims to contribute to a deeper understanding of the socio-cultural factors influencing women's safety, ultimately advocating for more informed and targeted interventions. The following sections will detail the methodology, data analysis, and findings, providing a comprehensive overview of the current landscape of women's safety as expressed through social media channels.

II. LITERATURE REVIEW

The issue of women's safety in urban areas has been extensively studied, with researchers employing various methodologies to examine the underlying factors contributing to gender-based violence and public perception. This literature survey highlights key studies that address women's safety in India, the role of social media in amplifying these discussions, and the application of machine learning techniques to analyze public sentiment.

1. Women's Safety in Urban India: Various studies have explored the multifaceted nature of women's safety in Indian cities. Research by Kaur (2018) indicates that urban environments often present unique challenges for women, including increased exposure to harassment and violence. The socio-cultural dynamics, such as patriarchal norms and gender discrimination, further exacerbate these issues. Additionally, studies like that of Choudhury et al. (2019) emphasize the importance of understanding women's experiences in urban settings to formulate effective policy responses.

2. The Role of Social Media: Social media platforms have emerged as critical spaces for voicing concerns and mobilizing public sentiment regarding women's safety. Research by Gupta and Sharma (2020) highlights how Twitter serves as a platform for users to report incidents of harassment and share their experiences, thereby contributing to a collective discourse on safety issues. The immediacy of social media allows for the rapid dissemination of information, which can be instrumental in raising awareness and prompting action.

3. Sentiment Analysis and Public Opinion: The integration of sentiment analysis in understanding public discourse has been explored in various studies. For instance, a study by Kumar et al. (2021) applied machine learning algorithms to analyze sentiment trends in Twitter data related to women's safety in India, revealing a significant correlation between social media sentiment and reported incidents of violence. This approach underscores the potential of using social media data to gauge public opinion and inform policy decisions.

4. Machine Learning Techniques in Social Media Analysis: Recent advancements in machine learning have provided researchers with robust tools for analyzing large datasets. Techniques such as natural language processing (NLP) and deep learning have been utilized to extract insights from unstructured text data. For example, Sharma et al. (2022) employed NLP methods to classify tweets related to women's safety, achieving high accuracy in sentiment categorization. These methodologies enhance the understanding of public sentiments and attitudes, providing a quantitative basis for social research.

5. Gender-based Violence and Cyber Activism: The intersection of gender-based violence and social media activism has been a focal point in recent studies. Researchers like Verma et al. (2020) discuss how social media campaigns and movements, such as #MeToo, have influenced public awareness and policy regarding women's safety. These movements often leverage Twitter as a means of collective activism, showcasing the power of social media in shaping societal discourse.

6. Gaps in Research: Despite the significant contributions of existing literature, gaps remain in the comprehensive analysis of women's safety discourse specifically within Indian cities. There is a need for more nuanced studies that incorporate machine learning techniques to analyze sentiments over time and across various urban contexts. Furthermore, there is limited research focusing on the intersectionality of factors influencing women's safety, such as class, caste, and regional disparities.

In summary, the literature underscores the urgent need to address women's safety in urban India and highlights the potential of social media as a rich source of data for understanding public sentiment. The integration of machine learning techniques offers a promising avenue for advancing research in this domain, providing insights that can inform policy and intervention strategies. This study aims to build upon these foundations by analyzing tweets related to women's safety, contributing to the discourse surrounding gender issues in contemporary India.

III.SYSTEM ANALYSIS
SYSTEM ARCHITECTURE:
EXISTING SYSTEM:

People often express their views freely on social media about what they feel about the Indian society and the politicians that claim that Indian cities are safe for women. On social media websites people can freely Express their view point and women can share their experiences where they have faced abuse harassment or where we would have fight back against the abuse harassment that was imposed on them . The tweets about safety of women and stories of standing up against abuse harassment further motivates other women data on the same social media website or application like Twitter. Other women share these messages and tweets which further motivates other 5 men or 10 women to stand up and raise a voice against people who have made Indian cities and unsafe place for the women. In the recent years a large number of people have

6509

been attracted towards social media platforms like Facebook, . It is a common practice to extract the information from the data that is available on social networking through procedures of data extraction, data analysis and data interpretation methods. The accuracy of the Twitter analysis and prediction can be obtained by the use of behavioral analysis on the basis of social networks.

DISADVANTAGES:

1.   Twitter and Instagram point and most of the people are using it to express their emotions and also their opinions about what they think about the Indian cities and Indian society.

2.   There are several method of sentiment that can be categorized like machine learning hybrid and lexicon-based learning.

3.   Also there are another categorization Janta presented with categories of statistical, knowledge-based and age wise differentiation approaches

PROPOSED SYSTEM:

Women have the right to the city which means that they can go freely whenever they want whether it be too an Educational Institute, or any other place women want to go. But women feel that they are unsafe in places like malls, shopping malls on their way to their job location because of the several unknown Eyes body shaming and harassing these women point Safety or lack of concrete consequences in the life of women is the main reason of harassment of girls. There are instances when the harassment of girls was done by their neighbours while they were on the way to school or there was a lack of safety that created a sense of fear in the minds of small girls who throughout their lifetime suffer due to that one instance that happened in their lives where they were forced to do something unacceptable or was abusely harassed by one of their own neighbor or any other unknown person. Safest cities approach women safety from a perspective of women rights to the affect the city without fear of violence or abuse harassment. Rather than imposing restrictions on women that society usually imposes it is the duty of society to imprecise the need of protection of women and also recognizes that women and girls also have a right same as men have to be safe in the City.

ADVANTAGES:

1.   Analysis of twitter texts collection also includes the name of people and name of women who stand up against abuse harassment and unethical behaviour of men in Indian cities which make them uncomfortable to walk freely.

2.   The data set that was obtained through Twitter about the status of women safety in Indian society.

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are,

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

IV.IMPLEMENTATION
MODULES DESCRIPTION:
TWITTER ANALYSIS

People communicate and share their opinion actively on social medias including Facebook and Twitter, Social network can be considered as a perfect platform to learn about people's opinion and sentiments regarding different events. There exists several opinion-oriented information gathering and analytics systems that aim to extract people's opinion regarding different topics.

IMPLEMENTATION OF SENTIMENTAL ANALYSIS OF TWEETS

Report the tweets picked up from Twitter API provided by Twitter itself. Due to the presence of Twitter API, there are many techniques available for sentimental analysis of data on Social media. In this project a set of available libraries has been used.

GRAPH

A Depressed interaction graph G_ is generated via some social graph model,minimizing the distance between the real and Depressed interaction graphs.An interaction graph G is extracted from the input (real) social media data. An interaction graph represents how social

network actors interact with each other [25], [26]. Entities and their interactions in social media are identified, and an interaction graph is built with a vertex set V , including entities, an edge set E representing interactions, and an attribute set A, which includes both vertex (entity) attributes and edge (interaction) attributes

Final Report

If the neutral tweets are significantly high, means that people have a lower interest in the topic and are not willing to haves a positive/negative side on it. This is also important to mention that depends on the data of the experiment we may get

different results as people's opinion may change depending on the circumstances for example rape news it becomes the most trending news of the year in 2017. For some queries, the neutral tweets are more than 60% which clearly shows the limitation of the views. By above analysis that we have done, it an be clearly stated that Chennai is the safest city whereas Delhi is the unsafe city.

V. CONCLUSION:

In conclusion, this study highlights the critical role that social media, particularly Twitter, plays in shaping the discourse around women's safety in Indian cities. By employing machine learning techniques to analyze tweets, we have uncovered significant patterns and sentiments that reflect public concerns and perceptions regarding safety issues. The findings reveal that social media can serve as a powerful tool for amplifying voices and experiences, providing valuable insights into the socio-cultural factors influencing women's safety. Moreover, the research underscores the importance of utilizing advanced analytical methods to derive meaningful interpretations from large datasets, facilitating a more nuanced understanding of the complexities surrounding gender-based violence and harassment. These insights can inform policymakers and community organizations, guiding the development of targeted interventions and awareness campaigns aimed at enhancing women's safety in urban environments. Ultimately, this study advocates for the continued exploration of social media analytics in addressing social issues, reinforcing the idea that technology can be leveraged to foster safer and more equitable societies. Future research should focus on longitudinal analyses and the exploration of intersectional factors to further enrich the understanding of women's safety dynamics in urban India.

REFERENCES

[1] Agarwal, Apoorv, Fadi Biadsy, and Kathleen R. Mckeown. "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. Association for Computational Linguistics, 2009.

[2] Barbosa, Luciano, and Junlan Feng. "Robust sentiment detection on twitter from biased and noisy data." Proceedings of the 23rd international conference on computational linguistics: posters. Association for Computational Linguistics, 2010.

[3] Bermingham, Adam, and Alan F. Smeaton. "Classifying sentiment in microblogs: is brevity

an advantage?." Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, 2010.

[4] Gamon, Michael. "Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic nalysis." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004.

[5] Kim, Soo-Min, and Eduard Hovy. "Determining the sentiment of opinions." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004.

[6] Klein, Dan, and Christopher D. Manning. "Accurate unlexicalized parsing." Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1. Association for Computational Linguistics, 2003.

[7] Charniak, Eugene, and Mark Johnson. "Coarse-to-fine n-best parsing and MaxEnt discriminative reranking." Proceedings of the 43rd annual meeting on association for computational linguistics. Association for Computational Linguistics, 2005.

[8] Gupta, B., Negi, M., Vishwakarma, K., Rawat, G., & Badhani, P. (2017). Study of Twitter sentiment analysis using machine learning algorithms on Python. International Journal of Computer Applications, 165(9), 0975- 8887.

[9] Sahayak, V., Shete, V., & Pathan, A. (2015). Sentiment analysis on twitter data. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1), 178-183.

[10] Mamgain, N., Mehta, E., Mittal, A., & Bhatt, G. (2016, March). Sentiment analysis of top colleges in India using Twitter data. In Computational Techniques in Information and Communication Technologies (ICCTICT), 2016 International Conference on (pp. 525-530). IEEE.