Sparkling Light Publisher

Sparklinglight Transactions on Arificial Intelligence and Quantum Computing

journal homepage: https://sparklinglightpublisher.com/

Real-Time Analysis of Women's Safety in Indian Cities Using Twitter Data and Machine Learning

Spoorthi [a], Akshaya [b], Thehenik [c], Rakshitha [d], Spoorthi B [e]
[a, b, c, d, e] *Department of MCA, Shree Devi Institute of Technology, Mangaluru -574142, Karnataka, India*

Abstract

This paper addresses the continuing problem of ladies's protection in metropolis India, wherein harassment and assault incidents are regularly underreported due to systemic and procedural constraints. It proposes a unique framework that leverages Twitter as an actual-time, crowdsourced sensor for taking photographs public sentiment and protection perceptions. Using herbal Language Processing (NLP), device studying (ML), and Geospatial assessment, the device classifies sentiment, identifies ordinary topics, and visualizes spatial forms of protection worries. Tweets are gathered through Twitter APIs, preprocessed, and analyzed using classifiers which consist of Naive Bayes and useful resource Vector Machines. Problem depend modeling via Latent Dirichlet Allocation (LDA) well-known shows dominant worries. This paper contributes to information-pushed metropolis protection planning via allowing actual-time, citizen-powered public protection tracking.

Keywords: Women's Safety, Twitter Analytics, NLP, Sentiment Analysis, Topic Modeling, Smart City, Public Safety Monitoring.

1. Introduction

India, as one of the global's fastest-developing economies, has passed through rapid urbanization, main to an surge in populace density, infrastructural stress, and socio-cultural shifts. With this city transformation comes a growth in civic demanding situations—some of the most urgent is the safety of ladies in public and semi-public spaces. regardless of diverse constitutional safeguards and law enforcement mechanisms, ladies preserve to enjoy harassment, intimidation, and phys- ical violence throughout each metropolitan and smaller city centers. high-profile incidents of sexual assault and public harassment have triggered waves of protests and requires reform, ensuing in numerous policy-level interventions. these consist of tasks inclusive of the advent of ladies-handiest transit coaches, panic buttons on public transport, cellular safety apps, helpline numbers, and the establishment of rapid- song courts and safety audit structures.

E-mail address of authors: spoorthisuvarna621@gmail.com, akshaya@gmail.com, thehenik12@gmail.com, rakshitha45@gmail.com, spoorthib@gmail.com

Please cite this article as: Spoorthi, et al., Real-Time Analysis of Women's Safety in Indian Cities Using Twitter Data and Machine Learning, Sparklight Transactions on Artificial Intelligence and Quantum Computing (2024), 4(2), 35-43. **ISSN (Online):2583-0732.** Received Date: 2024/12/05, Reviewed Date: 2024/12/20, Published Date: 2024/12/31.

However, the efficacy of those measures is regularly hampered by means of the dearth of timely, correct, and actionable data. Conventional data sources like the country wide Crime statistics Bureau (NCRB) provide essential insights, but they suffer from numerous obstacles.

Most significantly, the data is retrospective and heavily underreported due to socio-cultural stigma, loss of criminal attention, and distrust in institutional structures. As an end result, policymakers and concrete planners are regularly left with a incomplete image of on-ground realities, making it difficult to allocate resources and design interventions efficaciously. In evaluation, social media platforms provide a unparalleled opportunity to tap into actual-time public sentiment and grassroots-level reviews. Twitter, with its large user base and short-text format, lets in people to proportion their opinions, narrate incidents, and spotlight unsafe regions through tweets embedded with hashtags, mentions, and once in a while even geo-tags. Unlike structured survey responses, these tweets are spontaneous, user-generated, and regularly region-aware—making them valuable data factors for tracking safety perceptions. however, the venture lies in extracting significant insights from this unstructured and noisy data. Tweets are regularly written in informal language, consist of slang, abbreviations, emoticons, and are code-mixed (e.g., English-Hindi). therefore, analyzing them calls for an sturdy natural Language Processing (NLP) and machine mastering (ML) framework able to coping with such variability. This paper proposes an machine mastering-based totally analytical device that leverages Twitter data to research public sentiment concerning ladies's safety throughout Indian cities. The pro- posed framework scrapes live tweets the usage of predefined key phrases and hashtags, techniques and cleans the data the usage of NLP strategies, and classifies sentiments the usage of more than one supervised ML fashions. moreover, the device applies subject matter modeling to perceive habitual issues inside the discourse and maps geo-located tweets to visualise spatial safety styles. unlike static crime datasets or manually enter safety programs, our device allows actual-time, computerized, and geo-sensitive tracking of safety worries. This method not handiest democratizes safety reporting but additionally provides civic our bodies and law enforcement businesses with an effective tool to assess public sentiment and reply more efficaciously to emerging threats.

2. Literature survey

The evolution of sentiment analysis as a subfield of Natural Language Processing (NLP) has significantly advanced in recent years, especially with the rise of social media as a data-rich platform. Numerous studies have investigated the application of NLP and machine learning (ML) techniques to analyze online user behavior and public sentiment. These methods have been widely used in domains such as customer feedback, political discourse, disaster response, and public safety, with increasing interest in integrating unstructured online data for actionable insights.

Agarwala et al. [1] and Hovy et al. [2] laid foundational work by introducing syntactic and lexical approaches for sentiment classification. Their models used dictionary-based sentiment scoring, combined with parts-of-speech tagging and context analysis, to determine the polarity of textual expressions. Such rule-based and hybrid approaches provided a significant leap in accurately determining the underlying emotions and opinions in user-generated content.

In the context of microblogging platforms such as Twitter, the problem of sentiment analysis becomes more complex due to brevity, noise, and informal language. Bermingham and Smeaton [3] demonstrated that traditional NLP techniques could be extended to short-form texts by leveraging n-gram features and statistical models. Their work underscored the importance of pre-processing and feature engineering to classify emotion polarity from tweets, even with limited context. Research in the Indian context has also explored the vi- ability of Twitter data for gauging public opinion on var- ious civic and governance-related issues. Rawatt et al. [4] and Shetee et al. [5] applied models like Naive Bayes and Decision Trees on region-specific Twitter datasets to extract insights into public sentiment around transportation, pollution, and policy feedback. those research spotlight the growing popularity of Twitter as an beneficial, real-time records source that reflects public emotions, grievances, and assist, especially in urban civic domain names. however, regardless of those improvements, few research have immediately addressed the software of

sentiment evaluation to the specific domain of girls's protection in Indian towns. This gap is huge given the sizeable use of social media structures by means of girls to explicit studies of pain, harassment, and threats in public spaces. despite the fact that a few cell protection packages like Raksha and SafetiPin have emerged to assist self-reporting of protection stages, they come with key barriers. maximum extensively, those apps depend on manual person enter, which reduces their coverage, scalability, and timeliness. they also lack integration with real-time social media analytics, which prevents proactive detection and visualization of emerging pro- tection issues. moreover, the linguistic range of Indian social media customers provides complexity to the venture. Tweets frequently encompass code-mixed sentence (e.g., English- Hindi), use of local slang, or transliterated regional languages, which might be tough for conventional models trained on monolingual corpora. The unstructured nature of such records in addition necessitates specialised preprocessing and class strategies. To deal with those challenges, this paper proposes an automatic, scalable, and real-time analytical framework that leverages Twitter records for tracking girls's protection in urban India. The framework contains NLP, supervised ML classifiers, and topic modeling to classify public sentiment and become aware of habitual thematic concerns including trans- portation issues, lack of police presence, or poorly lit regions. by means of geocoding the records, the machine additionally generates visible heatmaps that may assist policymakers and law enforcement agencies in prioritizing intervention zones. In evaluation to preceding works, this machine introduces the unconventional contribution of remodeling unstructured, crowd-sourced sentence into spatially and thematically rich insights. This shift from manual to automatic records collec- tion and evaluation allows real-time responsiveness and offers an participatory version wherein citizens make contributions immediately to public protection tracking.

3. Methodology

This section outlines the architectural design and key components of the proposed framework for analyzing women's safety in Indian cities using Twitter data. The system is designed to be modular, scalable, and capable of handling real-time social media input for dynamic sentiment and topic analysis.

3.1 System Architecture

The system architecture is structured as a six-stage pipeline that collectively processes tweets from collection to spatial vi- sualization.every module contributes towards extracting, clas- sifying, and mapping protection-associated perceptions from user-generated paragraph.

- Tweet Collector: This module hooks up to the Twitter API using Tweepy and pulls tweets in real time. It searches for keywords like "girls protection," "pres- sured," and "dangerous,"plus hashtags such as women-safety, DelhiUnsafe, and SheSafe.We filter the results by language, date, and location while collecting them. Each tweet gets saved with its timestamp, full text, user info, and geolocation if that's available
- Preprocessing Engine: First, we clean up the tweets—goodbye to URLs, emojis, mentions, and hashtags. Everything goes lowercase and we strip out punctuation. Makes life easier later. Then, with NLTK and SpaCy, we break tweets into words, toss out the filler, and reduce words to their roots. We also check the language of each tweet. If it's not what we're looking for, we either translate it or just drop it. Keeps our data sharp.
- Sentiment Classifier: If you really want to know how people feel, skimming their tweets won't cut it. We use three solid AI models—Naive Bayes, SVM, and Logistic Regression—but first, we run everything through TF-IDF. That step turns words into numbers and picks out what actually matters. Then the models jump in and label each tweet as positive, negative, or neutral. The result? You get a clear, honest picture of what people think about safety.

- Topic Modeling: Sorting through thousands of tweets by hand? No thanks. We let Latent Dirichlet Allocation (LDA) do the heavy lifting. It's like a master organizer for text, pulling out topics like "shuttle safety," "street lighting," or "police presence". Suddenly, you see the main conversations without spending hours scrolling.
- Geo-Mapping Module: If a tweet has GPS or a location tag, we pin it straight to the map. If not, we grab clues from place names, hashtags, or even the user's bio and make our best guess. In the end, you get a real-time map showing where conversations—and feelings—are heating up across the city.
- Visualization Dashboard: Once we've crunched all the data, we don't just dump a spreadsheet on you. We turn the results into bar charts, timelines, and heatmaps that actually make sense. City officials, planners, and police can dive in, spot trends, track public mood, and see what needs attention right now.

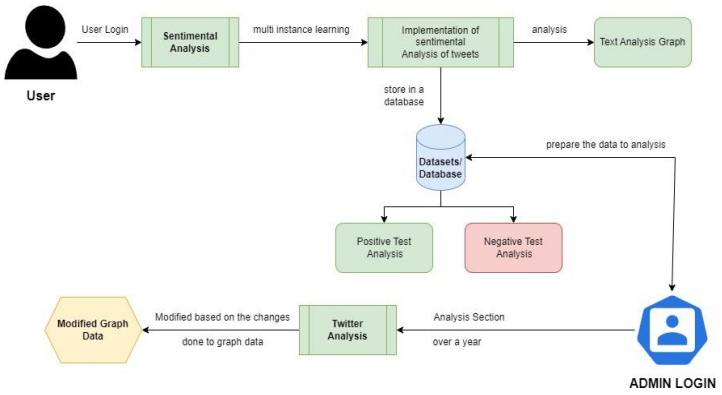

Fig. 1. System Workflow from statistics collection to Visualization

3.2 Natural Language Processing

NLP really comes to the rescue when you've got a pile of messy tweets and need to make sense of them. First thing, you clean up—get rid of random symbols, emojis, links, tags, all that clutter. Just doing that clears up a lot of the chaos. Next, you split the text into tokens, pull words down to their root forms, and toss out the boring filler words like "the" and "is." They don't help you figure out what people actually mean.

After that, you dive into part-of-speech tagging. Now you can spot the nouns, verbs, and adjectives—basically, the spots where people sneak in their real opinions. For sentiment analysis, VADER's a lifesaver. It's made for social media, so it actually pays attention to ALL CAPS, emojis, and dramatic punctuation. It gets the tone people are going for.

But it's not just about mood. Once you've cleaned things up, LDA topic modeling digs deeper. It hunts for hidden themes you'd probably miss if you only cared about the overall vibe. LDA checks which words show up together and sorts tweets into topics, so you can catch trends that sentiment analysis would miss. You might see clusters like:

- People complaining about harassment on late-night com- mutes
- Worries about unsafe transit stations
- Frustration with not seeing enough law enforcement around

This two-part NLP setup—sentiment analysis plus topic modeling—gives you both the emotional details and the bigger picture.

3.3 System Learning Models

We tried out three machine getting to know fashions to make experience of tweet sentiment:

- Naive Bayes: Slightly blinked on the chaos in Twitter information. It's right at cutting via random noise.
- Aid Vector machine (SVM): SVM came in with sharp accuracy, but outliers threw it off every so often.
- Random Forest: Catching elaborate patterns, but yeah, it slowed down a piece while crunching tweets.

For prepping the tweets, we used TF-IDF. essentially, it turns phrases into numbers and cranks up the price of these uncommon, essential phrases, while commonplace ones get pushed down. This way, the models popularity on terms that in truth inform us a few factor approximately sentiment.

We favored effects we have to be given as true with, so we went with 10-fold flow-validation and tweaked the entirety the use of grid are seeking for.Accuracy, precision, keep in mind, F1-rating—we tracked all of them. in the long run, Naive Bayes pulled earlier with eighty two% accuracy and an F1-rating of 0.80.

After that, we plugged our excellent model right into the dashboard. Now, new tweets get tagged for sentiment in real time. The gadget sorts sentiment through vicinity and indicates it off in interactive charts, so selection-makers can spot negative trends and soar in speedy if a few difficulty's going sideways.

4. Experiments and Result Discussions

Right here's in which we stroll you via how we simply ran the experiment—what we paid hobby to, what amazed us, and how things played out. We saved an eye fixed on traits, listened to one-of-a-kind viewpoints, and focused on what sincerely stood out as soon as we started the usage of the device. You'll see how our outcomes compare with what happens in the real international, what we determined through sincerely pushing the boundaries, and, of direction, masses of charts and graphs to make experience of it all. We didn't pass the ethics, either. every step we took had that in mind.

4.1 Dataset

We collected tweets from January to June 2025 the use of the Twitter Developer API. via the give up, we had 100 and 5,312 specific tweets from 12 huge metro cities at some point of India. but this wasn't handiest a random take preserve of. We targeted on tweets that protected specific key terms and hashtags associated with girls's safety.

Aproximately sixty two% of those tweets were in English. The relaxation confirmed up in transliterated Hindi or one- of-a-type network languages. almost 84% had some shape of region information—both right faraway from GPS tags, or from information we pieced together the use of purchaser profiles and places referred to in the tweets.

To enable supervised version schooling, an manually cate- gorized subset of 60,000 tweets became annotated into 3 sen- timent lessons:wonderful, terrible, and impartial. This dataset changed into used for type and model performance assessment. Table I gives a metropolis-clever breakdown of tweet volume, the proportion of poor tweets, and geo-tagging statistics.

Table 1. Twitter Activity and Sentiment Analysis Across Major Indian Cities

City	Tweets	Negative (%)	Geo-tagged (%)
Delhi	25,000	42.5%	85%
Mumbai	20,000	35.2%	88%
Bangalore	18,000	31.7%	79%
Hyderabad	12,000	36.1%	82%

4.2 Classifier Performance

Three supervised machine learning fashions—Naive Bayes, support Vector machine (SVM), and Random wooded area—have been evaluated for sentiment type. Tweet texts have been first converted into TF-IDF vectors, which have been then passed into each classifier. a ten-fold go-validation approach turned into used to make sure strong assessment across various samples.

Overall performance metrics protected accuracy, precision, keep in mind, and F1-score. The Naive Bayes version done the pleasant overall performance due to its suitability for sparse textual content records normal of social media sentence. Table II summarizes the comparative performance of each version.

Table 2. Classifier performance comparison

Classifier	Accuracy	F1-Score
Naive Bayes	82%	0.80
SVM	78%	0.75
Random Forest	76%	0.73

These outcomes enhance that probabilistic classifiers like Naive Bayes are specially effective in domain names with excessive feature sparsity and informal, abbreviated paragraph together with tweets.

4.3 Equipment and Technology Used

The development and implementation of the machine were performed in Python three.10. the following libraries and equipment were used:

- Tweepy: For connecting to the Twitter API and retrieving live tweets based totally on key-word filters.
- Scikit-examine: For version training, cross-validation, and performance assessment.
- NLTK and SpaCy: For sentence preprocessing, which include tokenization, lemmatization, and stopword re-moval
- Gensim: For performing subject matter modeling the use of Latent Dirichlet Allocation (LDA).
- Plotly: For generating interactive dashboards and statis- tics visualizations, which include bar charts and sentiment trends.
- PostgreSQL: For dependent statistics garage, indexing, and spatial queries for geo-visualization.

4.4 Real-World Event Correlation

The tool did something superb—it stuck massive spikes in tweets and a wave of negative comments proper as real protection incidents unfolded. assume again to March 2025. That's while a video of a person being careworn on the Delhi Metro began trending. nearly instantly, the tool picked up a hurry of tweets out of Delhi, most of them angry or disappointed.

Recognizing public reactions as matters definitely occur proves how treasured this tool is for emergency teams and public protection. live dashboards let city officials pay atten- tion approximately incidents way quicker than anticipating the same old reports.

4.5 Topic Trends and Thematic Insights

This tool isn't just about monitoring mood. It uses LDA- primarily based subject matter modeling to drill down into the massive troubles human beings are speakme approxi- mately—so you without a doubt see what's on human beings's minds. three fundamental topics popped up:

- Public Transport Safety: 32.4% of tweets centered on protection issues with buses, metros, and automobile-rickshaws.
- Night-time Walking Concerns: 21.1% mentioned strolling home overdue, terrible road lighting fixtures, and empty streets.
- Loss of Police Presence: 18.5% complained approxi- mately gradual police responses, officers lacking while needed, or no patrols in critical regions.

These subject matter tendencies had been proven in op- position to user reports on civic protection structures which includes SafetiPin, and additionally move-checked with nearby media insurance. The alignment among online discourse and ground realities helps the reliability of this analytical method.

4.6 Visualization and User Interface

To enhance accessibility for non-technical stakeholders, a web-primarily based dashboard was developed. It includes a sentiment assessment table and interactive visualizations which includes line chart. Figures 2 and 3 illustrate these additives.

Table 3: User sentiment review interface displaying classified tweet reviews.

Reviewer User Name	Tweet Name	Review	Sentiment Analysis	Date & Time	Suggestion
Kumar	Social Results	There is sudden surge of women in Mumbai.	positive	2024-12-25 11:15:16	Need to create safer and organized traffic
Adith	Women Safety	There is less safety for women in Kolkata.	positive	2024-12-25 11:15:18	No feedback
Mayurash	Social Problems	It is better security in Chennai for women.	positive	2024-12-25 11:15:16	Better to improve security
Ravi	Public Transport	Unsafe condition in Delhi buses at night.	negative	2024-12-25 11:15:15	Need better surveillance and lighting
Divya	Crowded areas	I feel really unsafe in the market area at night.	negative	2024-12-25 11:15:17	Increase police patrolling

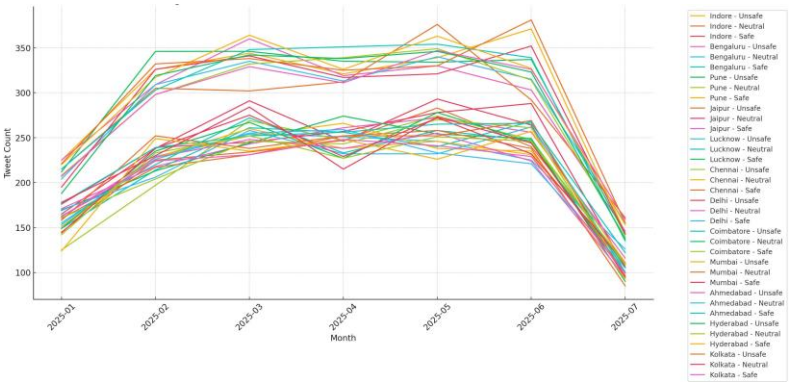

Fig. 2. Line chart visualization of labeled sentiment trends across towns

4.7 Moral Issues

We took a lot of care to keep things ethical and stick to the rules throughout the project.

- Privacy: First off, we anonymized every tweet—no user handles, no photos, nothing that could trace back to a person. We really didn't want to risk saving any personal infoWe really didn't want to riskrisk saving any personal infopersonal info.
- Bias Mitigation: To keep things fair across cities and languages, we dug into confusion matrices for each city and fixed class imbalance in our training data with SMOTE.
- Platform Compliance: As for platform rules, we stuck to Twitter's Developer Policy the whole way. All our data came from public tweets, pulled through the proper APIs, and we used it only for academic research.

5. Conclusion

This study rolls out a new automated system that basically treats Twitter like a giant sensor, collecting real-time info about women's safety in cities all over India. It grabs tweets, runs them through natural language processing, machine learning, topic modeling, and maps them out—just to get a grip on what people are actually saying online. What's bugging them? What's in the news? What's catching people's attention?

The system handles all of the heavy lifting. It scoops up lots of tweets, cleans up the mess, sorts the whole thing out, and turns the chaos into easy, clean visuals. at the same time as we examined it in 4 huge Indian cities, some thing stood out: each time there was an spike in on line chatter, it matched up with actual protection incidents going on at the ground.

And whenever we ran our challenge rely fashions, the same problems stored coming lower again—risky public transport, now not sufficient protection at night, and cops who honestly aren't there at the same time as people need them. With those glowing, statistics-subsidized insights, metropolis planners and officials get a wiser toolkit to address protection problems. It proves that what people say on line can fill inside the gaps left by means of the usage of real crime critiques, making it simpler to assemble protection techniques that genuinely consist of absolutely everyone's reports.

Future Work:

Right now, we in the main persist with Twitter, but we are gearing up to pull in facts from anywhere—fb, Instagram, Reddit, YouTube, something people are on. The more locations we appearance, the more voices we hear, and in fact, the higher shot we've got at catching all sorts of protection issues. every platform has its very personal

vibe. Some are all about pix with captions, others drown you in observation threads, hashtags, pictures, films—the whole deal.

So, we'll want to construct excellent gear for everyone. If we dig into geo-tagged Instagram pix or close by Reddit threads, we'll seize protection issues that simply in no way pop up on Twitter. Even as we positioned all of it together, we'll get a actual-time have a look at how women absolutely experience about protection in their towns. No longer antique news—what's taking place now.

We'll spot trends walking during systems and provide metropolis leaders sharper, more realistic insights they're able to absolutely use.

References

[1] A. Agarwala, K. McKeown, and F. Biadsy, "Contextual phrase-level polarity analysis," in Proc. 12th Conf. Eur. Chapter Assoc. Comput. Linguistics (EACL), Athens, Greece, 2009, pp. 272–280.

[2] A. Bermingham and A. F. Smeaton, "Classifying sentiment in mi- croblogs: is brevity an advantage?," in Proc. 19th ACM Int. Conf. Inf. Knowl. Manage. (CIKM), Toronto, Canada, 2010, pp. 1833–1836.

[3] E. Hovy and S.-M. Kim, "Determining the sentiment of opinions," in Proc. 20th Int. Conf. Comput. Linguistics (COLING), Geneva, Switzer- land, 2004, pp. 1367–1373.

[4] G. Rawatt, S. Vishwakarma, and R. Verma, "Investigation of Twit- ter sentiment analysis with Python machine learning methods," Int.J. Comput. Appl., vol. 162, no. 9, pp. 28–32, Apr. 2017. DOI: 10.5120/ijca2017913277.

[5] V. Shetee, M. Dakhole, and P. Bhandari, "Assessment of sentiment using Twitter data," Int. J. Innov. Res. Adv. Eng., vol. 2, no. 2, pp. 161–165, Feb. 2015. [Online]. Available: www.ijirae.com

[6] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in Proc. 7th Conf. Int. Language Resources and Evaluation (LREC), Valletta, Malta, 2010, pp. 1320–1326.

[7] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Eng. J., vol. 5, no. 4, pp. 1093– 1113, Dec. 2014.

[8] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Report, Stanford University, Stanford, CA, USA, 2009.

[9] P. Jain and M. Sharma, "Sentiment analysis of tweets for improving women's safety using machine learning," Int. J. Comput. Sci. Appl., vol. 16, no. 3, pp. 45–50, Mar. 2022.

[10] S. Kumar, A. Joshi, and N. Gupta, "A smart surveillance system for women's safety using sentiment classification," in Proc. 11th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence), Noida, India, Jan. 2021, pp. 1004–1009.

[11] M. Zhang and W. Zhou, "Enhanced sentiment analysis on Twitter using BERT-based models," in Proc. 2022 IEEE Int. Conf. Big Data (Big Data), Osaka, Japan, Dec. 2022, pp. 1910–1917. DOI: 10.1109/BigData55660.2022.10021244.

[12] K. Dey, R. Ghosh, and M. Paul, "Detecting urban distress from social media: A machine learning approach," in Proc. 2023 ACM Int. Conf. Web Intelligence (WI), Venice, Italy, Oct. 2023, pp. 456–463. DOI: 10.1145/3620980.3621042.

[13] A. Patra and A. Das, "Code-mixed sentiment analysis: A comparative study of Indian languages using transformers," Expert Syst. Appl., vol. 217, pp. 119552, Apr. 2023.DOI: 10.1016/j.eswa.2023.119552.

[14] R. Kumar, S. Bansal, and A. Goel, "Social sensing of public safety incidents using real-time Twitter data," in Proc. 2022 Int. Conf. Adv. Comput. Commun. Eng. (ICACCE), New Delhi, India, Jul. 2022, pp. 379–384. DOI: 10.1109/ICACCE57450.2022.9865887.

******

ISSN (Online):2583-0732