



**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

**Volume 9, Issue 12, December 2021**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.542**



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

# Analytical Survey on Twitter Data on Woman Safety Using Deep Learning Mechanism

Zahir Khan Shabbir Khan Mokashi, Ganesh Nitin Patil, Tousib Raju Bagwan, Arif Imtiaz Bargir,

Prof. Ashwini Kamble

Dept. of Computer, KJCOEMR, Pune, India

**ABSTRACT:** Safety of women has been one of the government's foremost critical demands, because it is one of essential prerequisites for a secure and satisfying lifestyle. However, it's far from the reality, even though there are numerous instances of harassment and assault towards women around the nation. The law enforcement authorities have already been working hard to secure the safety of women. Certain techniques have been extremely beneficial in minimizing incidents, but they have not particularly useful in boosting female confidence. This causes a great deal of anxiety and anguish for the women, which in itself is unwelcome and must be addressed. As a result, there is always an immediate need for a solution that may improve women's safety through an alarm system. A collection of studies on the issue of female rights were assessed for this aim. These approaches have been effective in achieving our approach that implements Bag of Words and Linear clustering along with machine learning algorithms such as Deep Belief Network and Decision Making. This approach will outlined further in the upcoming editions of this research.

**KEYWORDS:** Linear Clustering, Bag of Words, Deep Belief Networks.

## I. INTRODUCTION

In today's culture, safety of women is becoming one of the many pressing concerns, and despite the fact that perpetrators are being prosecuted harshly, the problem remains. Females are sometimes required to work irregular hours, yet they are compelled to leave their jobs by family and friends due to vulnerability. If females decide to work in several other regions, there will still be concerns about their family's protection and wellbeing. Female's personal security is a major worry in a variety of settings, particularly in emerging and underdeveloped nations. Numerous instances of violence and harassment, such as staring and passing remarks, are quite hostile, and these inappropriate activities are typically considered as a typical associated with urban lifestyle.

Numerous research has been undertaken in major Indian Cities, and women have experienced similar instances of sexual misconduct and receiving unwanted remarks by strangers. The fundamental cause for female misconduct is an absence of safety or real repercussions in the plight of women. There have been occasions in which girls have been sexually assaulted by their neighbors while walking home from school, or even where there has been a lack of security that has managed to create an atmosphere of anxiety in the thoughts of younger girls who have suffered throughout one's experience as a result of that one incident in which they were compelled to do anything inappropriate or were sexually exploited from one of their own neighbors or any other individual.

Women's safety in safe cities is approached from the standpoint of women's rights, allowing them to influence the city without fear of being attacked or gender discrimination. But instead of enforcing the limits that society generally puts on women, it is the responsibility of society to acknowledge the need for women's security and to realize that girls and women, like men, do have right to feel secure in the country. Despite how far technology has progressed, women's challenges remain unsolved. The current technology for women's protection offers a resolution to the dilemma, but part of it is large and non-wearable, and if it is, it may be easily recognized by an adversary, increasing the probability of the gadget being destroyed.

Social networking has been increasingly important in daily life in recent years. The percentage of people using social media is growing every day. Members share their opinions, feelings, and life events through social networking sites without restraint or reservation. Several social media platforms enable members to engage with only their acquaintances while still

allowing them to share their posts with a high degree of privacy. Users are migrating from conventional methods of communication including such blog posts or mailing lists to micro - blogging platforms such as Twitter, Facebook, and others due to convenient and straightforward privacy regulations and ease of accessibility to a certain social networks. Thousands of terabytes of words in the format of posts on social media makes it an appealing platform for academics to analyze data for the sake of women's issues.

There are indeed a variety of tactics and strategies, as well as the employment of legislation, to accomplish a practical and helpful increase in women's safety. Law enforcement authorities have already been working diligently to improve women's safety by employing multiple surveillance measures. Patrolling and camera transmissions are used by law enforcement authorities to keep an eye on the different locations. This is to guarantee that women are not assaulted and that the legal system is upheld. Even though these tactics have been beneficial in lowering crime, they have not been able to completely eradicate it. Safety of women is also jeopardized, and most females in poor nations are afraid to venture out alone.

The frequent occurrence of these assaults, as well as the amount of attack vectors that women are currently facing, creates a very dangerous situation for women. A mechanism that can adequately guarantee protection and safeguard the protection of women who need it most is considered essential. In this research study, a large number of studies have been examined for this goal. The assessment has been lucrative in terms of the successful implementation of our well-planned strategy. The suffering of women is a pressing concern, since the surge in harassment instances has resulted in a slew of inequities. The decline in women's safety creates a lot of suffering since they must be on heightened alert all or most of the time, which has resulted in a lot of psychological trauma for these females.

This literature survey paper segregates the section 2 for the evaluation of the past work in the configuration of a literature survey, and finally, section 3 provides the conclusion and the future work.

## II. RELATED WORKS

L. Wang et al. [1] examine sentiment dissemination on Twitter by examining a phenomenon called as sentiment reversal, and they identify numerous noteworthy elements of sentiment reversals utilizing repost cascade trees and repost diffusion networks. Following that, they develop SentiDiff, a breakthrough Twitter sentiment categorization system, as a sentiment reversal prediction model. SentiDiff addresses the interrelationships among textual content from Twitter tweets and sentiment diffusion patterns, and a supervised learning approach is utilized to combine the textual information-based sentiment classifier with the sentiment reversal prediction model. Experiments with real data show that the SentiDiff framework may help state-of-the-art textual information-based sentiment analysis algorithms achieve PR-AUC gains ranging from 5:09 to 8:38 percent.

B. S. Yaswanth et al. proposed a system in which depend on the user's location, the kNN algorithm is utilized to correctly forecast the nearest safe spot. Initially, safe place coordinates should be gathered and saved as a dataset within the system [2]. When the system is engaged, the algorithm at the back end does the various calculations based on the user's location. Because the dataset contains several safe place locations, the method should be chosen in such a manner that it can categorize the safe places and forecast the nearest place. Data categorization is needed by the standards. As a result, kNN is one of the most basic categorization methods. The data set includes safe locations such as a police station, a bus stop, or any populated area.

S. Hiremath et al. suggest an unsupervised sentiment classification using emoticons on Twitter Data. The labeled data is not necessary because the approach is unsupervised. Because the Lexicon is produced depending on the Emoticons, the classifier may be used to categorize tweets from any domain [3]. Tweets with emoticons are initially utilized to train the classifier. The sentiment words are categorized as positive, negative, or neutral depending on the sentiment of the emoticon in the tweet. Once the classifier has been trained, tweets without emoticons are classified using the lexicon that was produced while the classifier was being trained. The outcome shows that unsupervised classification with Emoticons outperformed semi-supervised classification without Emoticons.

D. Kumar investigated numerous machine learning approaches that can help organize and analyze the huge amount of Twitter data collected, including millions of tweets and text messages published every day. These machine learning approaches, including the SPC algorithm and linear algebraic factor model techniques, are very strong and useful when it comes to analyzing huge volumes of data, and they help to further categorize the data into meaningful categories [4]. The dataset on the state of women's safety in Indian society that was collected through Twitter was processed utilizing machine learning algorithms for the intent of smoothing the data by eliminating zero values and developing a data analysis technique by utilizing Laplace and Porter's theory to eliminate re-tweet and redundant data from the data set that was received so that a clear and original perspective of women's safety in Indian society could be achieved.

R. Rani discovered that sentiment analysis is the most effective method for assessing positive, negative, and neural data from tweets. The classification approach may be used to do sentiment analysis. Categorization methods can separate data into groups. In this study, the SVM and KNN classifications are applied [5]. The SVM is a classification approach that identifies data classes using a single hyperplane. The KNN is a classification method that uses the Euclidian distance to define several hyperplanes. The proposed framework outperforms the SVM classifier in terms of accuracy when the new and current models are implemented in Anaconda.

Using real Twitter data, A. Aslam et al. proposed an architecture that provides users with extensive perspectives on any topic of interest. It provides the findings in both graph and text form for deep extraction of user opinions. The most recent and popular thoughts on any subject, product, celebrity, organization, initiative, or movement run by the government are retrieved and shown to the party of interest. It assists educational institutions in assessing student behavior and activities [6]. It also plainly assists candidates during elections by providing them with people's expectations and concerns, as well as their geographical region, it also helps marketers and companies enhance their standards and sales appropriately. Recommendations are also made based on previous searches.

G. A. J. Satvika et al. developed a system that was capable of determining congestion spots in the Jakarta region by leveraging data from Twitter and providing the best route recommendations for users [7]. This system employs the KNearest Neighbor (KNN) approach, which is designed to classify or process data mining to determine whether or not the classification results are jammed. An assessment was performed in this research to determine the accuracy, precision, and recall outcomes of computations using the K-Nearest Neighbor approach. It is believed that by employing this strategy, excellent performance in categorizing data, such as high accuracy, will be obtained.

L. G. S. Selvan creates a more effective structure for firms to track their products. It allows businesses to see how customers feel about their products and monitor bugs. The presented approach takes advantage of a real-time data stream from Twitter. These data streams are filtered and evaluated, and quick response is acquired via opinion mining [8]. To deal with the massive amount of data transmitted from Twitter, the framework depends on Apache Hadoop. The trials demonstrated an 84 percent accuracy in emotional analysis.

A. P. Jain proposes a technique for predicting the general feelings and tendency of Indians toward political situations and issues. Twitter API v 1.1 is used to obtain raw training tweets. Following the collection of raw tweets, several preprocessing methods are used to clean the data. The same methods are used to gather and sanitize raw tweets to prepare the testing dataset [9]. Following the production of training and testing datasets, several classifiers are utilized to examine classifier performance.

Using tweet data acquired from social media, Twitter, crawlers, M. Wongkar [10] discusses the public's mood about the Republic of Indonesia's presidential contenders for the 2019-2024 timeframe. The author also executes text processing with the data and applies the Naive Bayes framework to forecast the class. He divides the world into two categories: positive and negative. According to the findings of trials, the Naive Bayes approach has higher accuracy than other methods, such as KNN, which has an accuracy rate of 75.58 %, and SVM, which has an accuracy rate of 63.99 %.

M. R. Hasan et al. created an NLP-based preprocessed data framework to filter tweets using the Bag of Word (BoW) model and the TF-IDF model for sentiment analysis. It has performed tokenization, stemming, stopword removal, named entity identification, and text modeling as Bag of Word and TF-IDF Model [11]. The main goal is to determine the emotion of a tweet by defining positive and negative polarity, and tweets are gathered using Twitter's streaming API. The

authors use these tweets as raw data. At that moment, they employ the provided approach for tweet evaluation. Before making a purchasing decision, the client will comprehend the feedback on the services depending on the sentiment analysis.

D. Goularas tested the deep learning algorithms depending on CNN and LSTM networks in various configurations for sentiment analysis in Twitter data. This evaluation produced somewhat poorer but comparable findings to state-of-the-art approaches, allowing solid conclusions about the various configurations to be drawn. The limited performance of these systems demonstrated the limits of CNN and LSTM networks in practice. In terms of setup, it was discovered that when CNN and LSTM networks are coupled, they perform better than when used alone [12]. This is because CNNs have an excellent dimensionality reduction technique while LSTM networks preserve word dependencies. The difference in accuracy performance across different datasets illustrates that, as predicted, having an adequate dataset is the most important factor in improving the performance of such systems.

V. Rekha et al. explain that utilizing hashtags to extract tweets based on a topic of interest proved to be effective. The data was extracted in comma-separated values and cleaned, including the elimination of unnecessary columns, expressions, punctuation, and so on. The removal of stop words aided in the calculation of accuracy. It was discovered that stemming did not result in a substantial improvement in sentiment computation. The cleansed data was fed through three machine learning classifiers: naive bayes, random forest, and customized random forest [13]. Random forest outperformed naive bayes in sentiment prediction accuracy. Tailored random forest entailed the addition of a few new parameters, which resulted in a noticeable increase in prediction accuracy. These forecasts can be used by the government to gauge public sentiment on various projects that have been implemented.

L. Mandloi addressed many machine learning tools for data analysis on Twitter, such as Naive Bayes, SVM, and Maximum Entropy technique. Twitter data is being analyzed in a variety of ways to harvest sentiments. This research describes the concept of opinion in Twitter sentiment analysis [14]. Sentiment analysis is concerned with opinions that are characterized as favorable, negative, or neutral. According to the study, machine learning tools such as Naive Bayes have the best accuracy and may be considered as baseline learning methods, although Maximum Entropy approaches are also quite useful in specific circumstances.

The notion of fine-grained sentiment analysis was created by S. Tiwari et al. using one of the most popular social networking sites, "Twitter", which is well-known and widely utilized throughout the world [15]. The assessment of feature extraction from Twitter data has been finished and may be utilized for further analysis and decision-making to mine the opinion or sentiment. For the current investigation, it is also discovered that the decision tree and random force algorithms are more accurate than the SVM approach.

### III. CONCLUSION AND FUTURESCOPE

One of the most pressing needs of the era is the development of a trustworthy women's rights monitoring. The number of assaults on women, such as harassment as well as other forms of assault, has grown. These really are bad because they have enabled societies to erode and caused females in emerging economies to look over their shoulders. This has caused a lot of anxiety and anguish for women, which has resulted in hundreds of issues in their lives. This also has a negative impact on women's standard of living as well as their health and wellbeing. As a result, an implementation plan for women's safety warning is required, and this survey study has examined several research publications in depth. This has allowed us to reach the methodology for such a system through the use of machine learning algorithms such as Deep Belief Networks and Decision Making. This methodology will be elaborated further in the upcoming editions of this research.

### REFERENCES

- [1] L. Wang, J. Niu and S. Yu, "SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis," in IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 10, pp. 2026-2039, 1 Oct. 2020, DOI: 10.1109/TKDE.2019.2913641.
- [2] B. S. Yaswanth, R. S. Darshan, H. Pavan, D. B. Srinivasa, and B. T. V. Murthy, "Smart Safety and Security Solution for Women using kNN Algorithm and IoT," 2020 Third International Conference on Multimedia Processing, Communication & Information Technology (MPCIT), 2020, pp. 87-92, DOI: 10.1109/MPCIT51588.2020.9350431.



- [3] S. Hiremath, S. H. Manjula and V. K. R, "Unsupervised Sentiment Classification of Twitter Data using Emoticons," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 444-448, DOI: 10.1109/ESCI50559.2021.9397026.
- [4] D. Kumar and S. Aggarwal, "Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets," 2019 Amity International Conference on Artificial Intelligence (AICAI), 2019, pp. 159-162, DOI: 10.1109/AICAI.2019.8701247.
- [5] R. Rani and S. Tandon, "Chat Summarization and Sentiment Analysis Techniques in Data Mining," 2018 4th International Conference on Computing Sciences (ICCS), 2018, pp. 102-106, DOI: 10.1109/ICCS.2018.00025.
- [6] A. Aslam, U. Qamar, R. A. Khan, P. Saqib, A. Ahmad, and A. Qadeer, "Opinion Mining Using Live Twitter Data," 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), 2019, pp. 36-39, DOI: 10.1109/CSE/EUC.2019.00016.
- [7] G. A. J. Satvika, S. M. Nasution, and R. A. Nugrahaeni, "Determination of the Best Vehicle Pathway with Classification of Data Mining Twitter using K-Nearest Neighbor," 2018 International Conference on Information Technology Systems and Innovation (ICITSI), 2018, pp. 72-76, DOI: 10.1109/ICITSI.2018.8695947.
- [8] L. G. S. Selvan and T. Moh, "A framework for fast-feedback opinion mining on Twitter data streams," 2015 International Conference on Collaboration Technologies and Systems (CTS), 2015, pp. 314-318, DOI: 10.1109/CTS.2015.7210440.
- [9] A. P. Jain and V. D. Katkar, "Sentiments analysis of Twitter data using data mining," 2015 International Conference on Information Processing (ICIP), 2015, pp. 807-810, DOI: 10.1109/INFOP.2015.7489492.
- [10] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), 2019, pp. 1-5, DOI: 10.1109/ICIC47613.2019.8985884.
- [11] M. R. Hasan, M. Maliha, and M. Arifuzzaman, "Sentiment Analysis with NLP on Twitter Data," 2019 International Conference on Computer, Communication, Chemical, Materials, and Electronic Engineering (IC4ME2), 2019, pp. 1-4, DOI: 10.1109/IC4ME247184.2019.9036670.
- [12] D. Goularas and S. Kamis, "Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data," 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), 2019, pp. 12-17, DOI: 10.1109/Deep-ML.2019.00011.
- [13] V. Rekha, R. Raksha, P. Patil, N. Swaras and G. L. Rajat, "Sentiment Analysis on Indian Government Schemes Using Twitter data," 2019 International Conference on Data Science and Communication (IconDSC), 2019, pp. 1-5, DOI: 10.1109/IconDSC.2019.8817036.
- [14] L. Mandloi and R. Patel, "Twitter Sentiments Analysis Using Machine Learning Methods," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-5, DOI: 10.1109/INCET49848.2020.9154183.
- [15] S. Tiwari, A. Verma, P. Garg, and D. Bansal, "Social Media Sentiment Analysis On Twitter Datasets," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, pp. 925-927, DOI: 10.1109/ICACCS48705.2020.9074208.



**INNO**  **SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 7.542**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
**INDIA**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details