



---

## **Analysis of Women Safety using Machine Learning Techniques on Tweets**

*<sup>1</sup>Pabboju Archana, <sup>2</sup>Dr. V. Bapuji*

*<sup>1,2</sup>Vaageshwari College of Engineering, Karimnagar*

---

### **ABSTRACT**

Women and girls have been experiencing a lot of violence and harassment in public places in various cities starting from stalking and leading to sexual harassment or sexual assault. This research paper basically focuses on the role of social media in promoting the safety of women in Indian cities with special reference to the role of social media websites and applications including Twitter platform Facebook and Instagram. This paper also focuses on how a sense of responsibility on part of Indian society can be developed the common Indian people so that we should focus on the safety of women surrounding them. Tweets on Twitter which usually contains images and text and also written messages and quotes which focus on the safety of women in Indian cities can be used to read a message amongst the Indian Youth Culture and educate people to take strict action and punish those who harass the women. Twitter and other Twitter handles which include hash tag messages that are widely spread across the whole globe sir as a platform for women to express their views about how they feel while we go out for work or travel in a public transport and what is the state of their mind when they are surrounded by unknown men and whether these women feels safe or not?

Keywords: Machine learning, Sexual Harrasment, Sentimental analysis, Safety, Women

---

### **Introduction**

Twitter in this modern era has emerged as a ultimate microblogging social network consisting over hundred million users and generate over five hundred million messages known as 'Tweets' every day. Twitter with such a massive audience has magnetized users to emit their perspective and judgemental about every existing issue and topic of internet, therefore twitter is an informative source for all the zones like institutions, companies and organizations. On the twitter, users will share their opinions and perspective in the tweets section. This tweet can only contain 140 characters, thus making the users to compact their messages with the help of abbreviations, slang, shot forms, emoticons, etc. In addition to this, many people express their opinions by using polysemy and sarcasm also. Hence twitter language can be termed as the unstructured. From the tweet, the sentiment behind the message is extracted. This extraction is done by using the sentimental analysis procedure. Results of the sentimental analysis can be used in many areas like sentiments regarding a particular brand or release of a product, analyzing public opinions on the government policies, people thoughts on women, etc. In order to perform classification of tweets and analyze the outcome, a lot of study has been done on the data obtained by the twitter. We also review some studies on machine learning in this paper and research on how to perform sentimental analysis using that domain on twitter data. The paper scope is restricted to machine learning algorithm and models. Staring at women and passing comments can be certain types of violence and harassments and these practices, which are unacceptable, are usually normal especially on the part of urban life. Many researches that have been conducted in India shows that women have reported sexual harassment and other practices as stated above. Such studies have also shown that in popular metropolitan cities like Delhi, Pune, Chennai and Mumbai, most women feel they are unsafe when surrounded by unknown people. On social media, people can freely express what they feel about the Indian politics, society and many other thoughts. Similarly, women can also share their experiences if they have faced any violence or sexual harassment and this brings innocent people together in order to stand up against such incidents. From the analysis of tweets text collection obtained by the twitter, it includes names of people who has harassed the women and also names of women or innocent people who have stood against such violent acts or unethical behaviour of men and thus making them uncomfortable to walk freely in public.

---

### **Existing work**

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 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.

#### **How machine learning is used for Tweets:**

**Machine learning algorithms constantly run in the background, analyzing tweets and engagement data. Based on this data, the algorithm ranks tweets and puts them in order of most relevant to least relevant.**

This ensures that users see the tweets that are most likely to be interesting or engaging to them. Additionally, Twitter uses AI to identify bots and spam accounts. This helps to keep the platform clean and prevents users from seeing irrelevant or unwanted content. Using these technologies, Twitter can provide users with a better experience by showing them the best tweets first.

#### **Tweets:**

#### **Robust sentiment detection on twitter from biased and noisy data.**

we propose an approach to automatically detect sentiments on Twitter messages (tweets) that explores some characteristics of how tweets are written and meta-information of the words that compose these messages. Moreover, we leverage sources of noisy labels as our training data. These noisy labels were provided by a few sentiment detection websites over twitter data. In our experiments, we show that since our features are able to capture a more abstract representation of tweets, our solution is more effective than previous ones and also more robust regarding biased and noisy data, which is the kind of data provided by these sources

---

#### **Present work**

we have discussed about various machine learning algorithms that can help us to organize and analyze the huge amount of Twitter data obtained including millions of tweets and text messages shared every day. We have used three types of machines. Throughout the research paper, we have discussed about various machine learning algorithms that can help us to organize and analyze the huge amount of Twitter data obtained including millions of tweets and text messages shared every day. We have used three types of machine learning algorithms in this paper.

1. Random Forest Classifier.
2. Logistic Regression.
3. Decision Tree Classifier.

Random forest classifier is a popular ML algorithm that comes under supervised learning technique. It is used for classification in ML. Logistic regression algorithm is based on the concept of probability and its cost function is between 0 and 1. Decision tree classifier is used to continuously split the data according to a certain parameter. These machine learning algorithms are highly effective and useful when it comes to analyzing of large amount of data including the SPC algorithm and linear algebraic Factor Model approaches which help to further categorize the data into meaningful groups. Support vector machine is yet another form on machine learning algorithm that is extremely popular in extracting. Useful information from the Twitter and get an idea about the status of women safety in Cities.

---

#### **CONCLUSION**

Two dynamic searchable encryption techniques with a high level of security were provided by the researchers. Both forward and backward collusion between the cloud server and search users can be prevented by the first method. Another solution to the problem of key sharing in the kNN-based searchable encryption system is provided by the second. According to performance assessments, the new methods outperform the existing ones in terms of storage, search, and update complexity. Testing shows that schemes are efficient in terms of storage overhead, index construction and trapdoor creation and queries.

---

#### **Plans for the Future**

Talking about the future work, we can implement tracking system by asking every woman to switch on the location button in her mobile. And that should be connected to her parents as well as the nearest police station. So, that they can easily know the location and will be able to help her. We should use this Machine Learning algorithm on other social media platforms like Facebook, Instagram etc. because in our project we have used only twitter. So, that we can apply this on large scale and help women in larger scale helping to build a happier and healthy nation.

---

#### **References**

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.

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

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.

- 
- Gamon, Michael. "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. Association for Computational Linguistics, 2004.
- 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.
- J.-J. Wang, N.-N. Zhao, and J.-H. Li, "Current situation of marine microplastics pollution and prevention proposals in China," China Environmental Science, vol. 39, no. 7, pp. 3056-3063, Jul. 2019, doi:10.19674/j.cnki.issn1000-6923.2019.0360.
- W.-B. Li, G. Ma, E.-Q. Yang, Y.-M. Cai, Z. Chen, R.-f. Gao, J.-H. Yan, X.-F. Cao, and E.-J. Pan, "Study on characteristics of electric dust removal fly ash and bag fly ash in circulating fluidized bed waste incineration system," Proceedings of the CSEE, vol.39, no.5, pp.1397-1405, Mar.2019, doi: 10.13334/j.0258-8013.pcsee.181110.
- D. Porshnov, V. Ozols, and M. Klavins, "Thermogravimetric analysis as express tool for quality assessment of refuse derived fuels used for pyro-gasification," Environmental Technology, vol. 41, no. 1, pp. 39-35, Mar 2020.
- P. Kellow, R. Joel J P C, D. Ousmane, D. Ashok Kumar, d.- A. Victor Hugo C, and K. Sergei A, "A Smart Waste Management Solution Geared towards Citizens," Sensors (Basel, Switzerland), vol. 20, no. 8, pp.1-15, Apr.2020, doi:10.3390/s20082380.
- Y.-G. Cheng, N. Chen, and H. Zhang, "Coal Fly Ash as an Inducer to Study Its Application in the Production of Methane Gas from Domestic Waste," Fresenius Environmental Bulletin, vol. 29, no. 2, pp. 1082-1089, 2020.