

**A REVIEW ON ANALYSIS OF WOMEN SAFETY IN INDIAN CITIES USING MACHINE LEARNING ON TWEETS**

**KODURU TEJA SRI DEVI<sup>1</sup> MOHAMMED SHAHEENA SULTHANA<sup>2</sup> GARIKAPATI SAI SASI KUMAR<sup>3</sup>  
RAYAPATI.KANAKA DURGA<sup>4</sup>**

<sup>1</sup> M.C.A Student, Dept of M.C.A, Amrita Sai Institute of Science & Technology, Paritala, A.P., India

<sup>2</sup> M.C.A Student, Dept of M.C.A, Amrita Sai Institute of Science & Technology, Paritala, A.P., India

<sup>3</sup> M.C.A Student, Dept of M.C.A, Amrita Sai Institute of Science & Technology, Paritala, A.P., India

<sup>4</sup> Assistant Professor, Dept of M.C.A, Amrita Sai Institute of Science & Technology, Paritala, A.P., India

**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. 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 feel safe or not?

Keywords: Women, Safety, Sexual Harassment, Hash tag, Sentimental Analysis.

**INTRODUCTION**

There are certain types of harassment and Violence that are very aggressive including staring and passing comments and these unacceptable practices are usually seen as a normal part of the urban life. There have been several studies that have been conducted in cities across India and women report similar type of sexual harassment and passing off comments by other unknown people. The study that was conducted across most popular Metropolitan cities of India including Delhi, Mumbai and Pune, it was shown that 60 % of the women feel unsafe while going out to work or while travelling in public transport.

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 sexually 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 sexual 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. Analysis of twitter texts collection also includes the name of people and name of women who stand up against sexual harassment and unethical behaviour of men in Indian cities which make them uncomfortable to walk freely. The data set that was obtained through Twitter about the status of women safety in Indian society was for the processed through machine learning algorithms for the purpose of smoothening the data by removing zero values and using Laplace and porter's theory is to developer method of analyzation of data and remove retweet and redundant data from the data set that is obtained so that a clear and original view of safety status of women in Indian society is obtained..

## **LITERATURE SURVEY**

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 [1]. On social media websites people can freely express their view point and women can share their experiences where they have faced sexual harassment or where we would have fight back against the sexual harassment that was imposed on them [2]. The tweets about safety of women and stories of standing up against sexual 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, 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.

There are several method of sentiment that can be categorized like machine learning hybrid and lexicon-based learning. [5] Also there are another categorization Janta presented with categories of statistical, knowledge-based and age wise differentiation approaches. 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 behavioural analysis on the basis of social networks.

## **SYSTEM ANALYSIS**

The Systems Development Life Cycle (SDLC), or Software Development Life Cycle in systems engineering, information systems and software engineering, is the process of creating or altering systems, and the models and methodologies that people use to develop these systems. In software engineering the SDLC concept underpins many kinds of software development methodologies.

## **EXISTING SYSTEM**

This Concept to analyse women safety using social networking messages and by applying machine learning algorithms on it. Now-a-days almost all peoples are using social networking sites to express their feelings and if any women feel unsafe in any area then she will express negative words in her post/tweets/messages and by analysing those messages we can detect which area is more unsafe for women's.

## **DISADVANTAGES**

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

## **PROPOSED SYSTEM**

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.

In propose work author using TWEETPY package from python to download tweets from twitter but every time INTERNET will not available to download tweets online so we downloaded MEETOO tweets on women safety and safe inside dataset folder. Application will read the tweets to detect women's sentiments.

Author using NLTK (natural language tool kit) to remove special symbols and stop words from tweets and to make them clean

## **ADVANTAGES**

Author using TEXTBLOB corpora package and dictionary to count positive, negative and neutral polarity and the tweets which has polarity value less than 0 will consider as negative as and greater than 0 and less than 0.5 will consider as neutral and polarity greater than 0.5 will consider as positive.

## **DATASET**

MeeToo\_tweets.csv

## **ALGORITHM**

Sentimental analysis algorithm

## **REQUIREMENTS SPECIFICATION**

Requirement Specification provides a high secure storage to the web server efficiently. Software requirements deal with software and hardware resources that need to be installed on a serve which provides optimal functioning for the application. These software and hardware requirements need to be installed before the packages are installed. These are the most common set of requirements defined by any operation system. These software and hardware requirements provide a compatible support to the operation system in developing an application.

## **HARDWARE REQUIREMENTS:**

The hardware requirement specifies each interface of the software elements and the hardware elements of the system. These hardware requirements include configuration characteristics.

- System : Pentium IV 2.4 GHz.

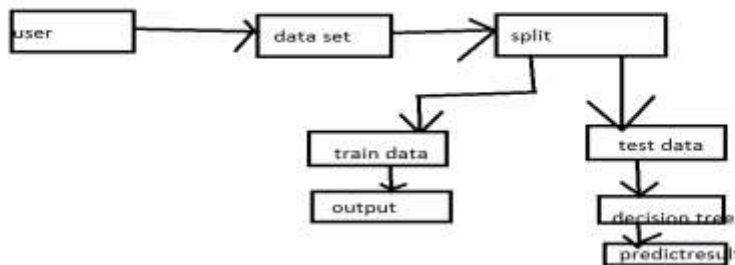
- Hard Disk : 100 GB.
- Monitor : 15 VGA Color.
- Mouse : Logitech.
- RAM : 1 GB.

### **SOFTWARE REQUIREMENTS:**

The software requirements specify the use of all required software products like data management system. The required software product specifies the numbers and version. Each interface specifies the purpose of the interfacing software as related to this software product.

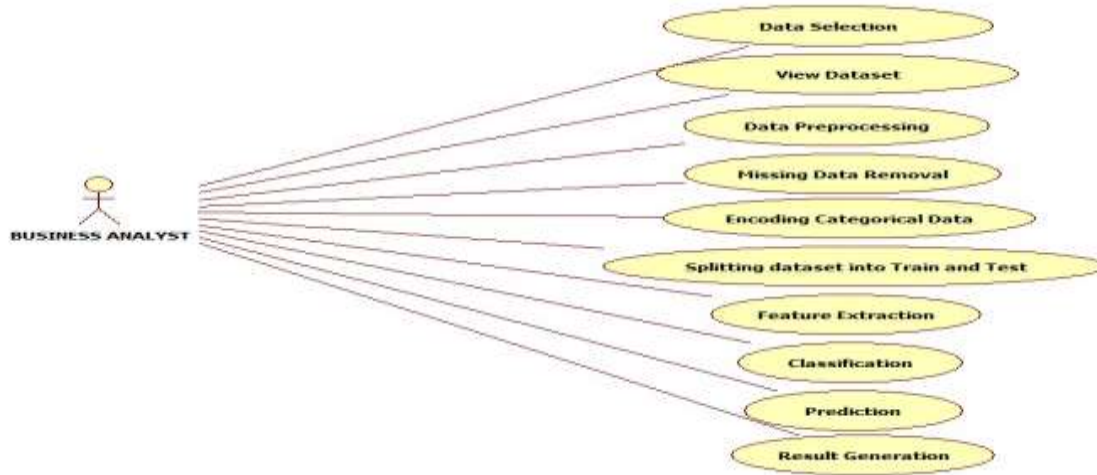
- Operating system : Windows XP/7/10
- Coding Language: Python 3.7

### **SYSTEM ARCHITECTURE**

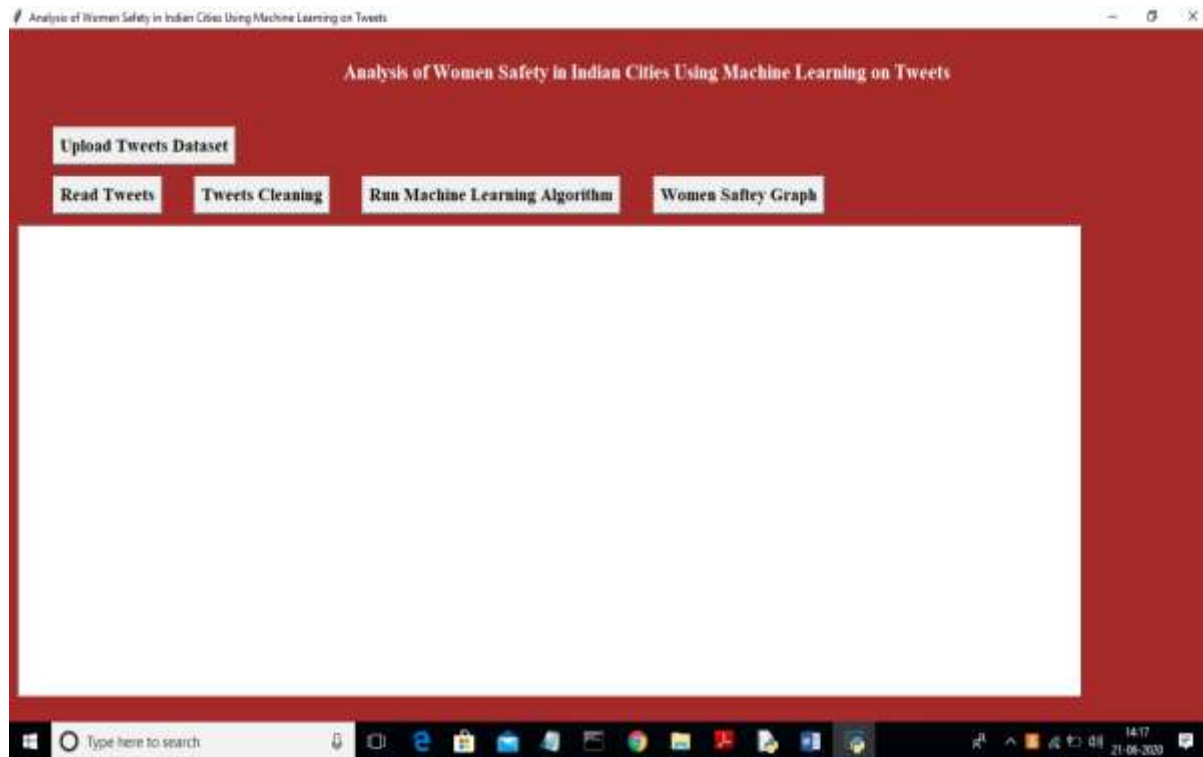


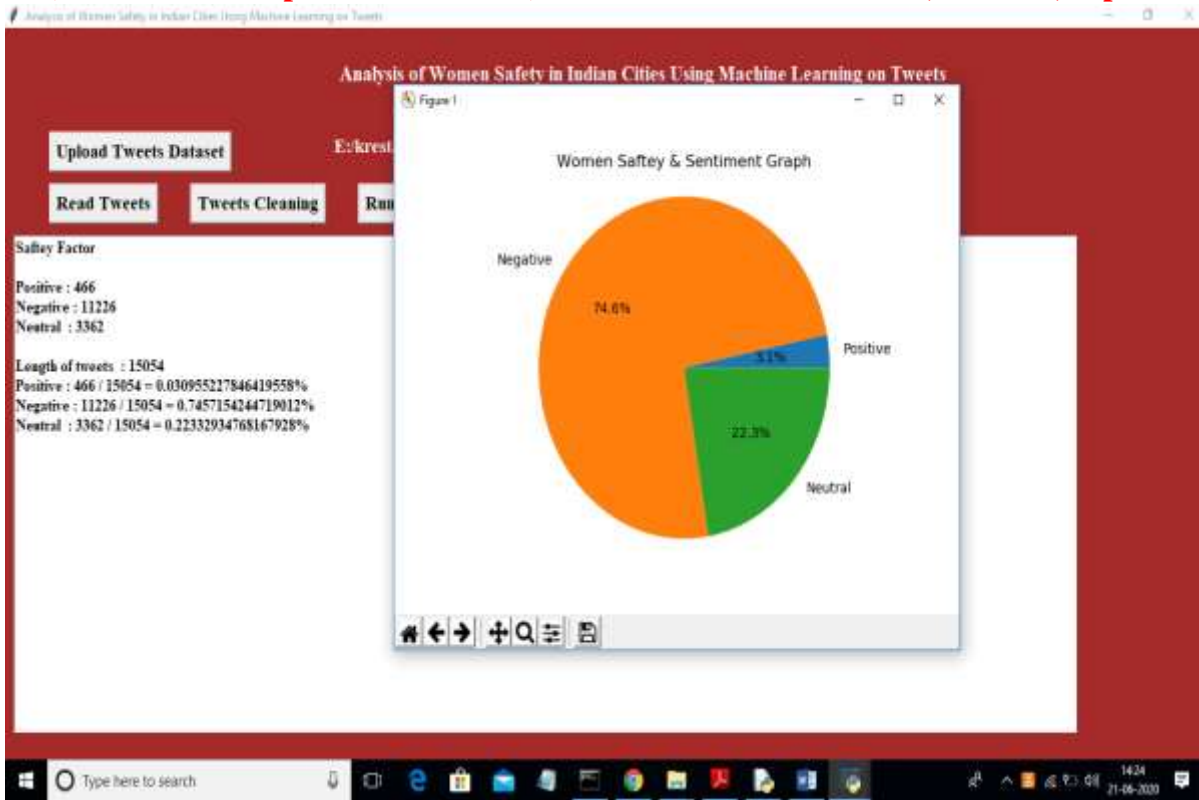
### **USE CASE DIAGRAM**

A use case diagram at its simplest is a representation of a user's interaction with the system and depicting the specifications of a use case. A use case diagram can portray the different types of users of a system and the various ways that they interact with the system. This type of diagram is typically used in conjunction with the textual use case and will often be accompanied by other types of diagrams as well.



### SCREEN SHOTS





## CONCLUSION

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. These machine learning algorithms are very 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 machines is yet another form of machine learning algorithm that is very popular in extracting Useful information from the Twitter and get an idea about the status of women safety in Indian cities..

## FUTURE ENHANCEMENTS:

It is not possible to develop a system that makes all the requirements of the user. User requirements keep changing as the system is being used. Some of the future enhancements that can be done to this system are:

- As the technology emerges, it is possible to upgrade the system and can be adaptable to desired environment.
- Based on the future security issues, security can be improved using emerging technologies like single sign-on.

## **REFERENCES**

- [1] Agarwal, Apoorv, Fadi Biadys, 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 analysis." 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