International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 02 | Feb 2023 www.irjet.net p-ISSN: 2395-0072

Analysis Of Tweets Using Machine Learning to Examine Women's Safety in Indian Cities

Ashwini Y[1], Kavya M[2], Deekshitha A [3], Akshitha [4], Bhavya Balakrishnan[5]

[1,2,3,4]*Students, Department of Computer Science & Engineering, T John Institute of Technology, Bangaluru, India*
[5]*Asst. Professor, Department of Computer Science & Engineering, T John Institute of Technology, Bangaluru, India*

Abstract - *In many cities, violence and harassment against women and girls in public spaces has increased, starting with stalking and progressing to sex harassment or sexual assault. The primary topic of this study paper is on how social media, specifically the Twitter platform, Facebook, and Instagram, plays a part in enhancing the safety of women in Indian cities. This essay also focuses on how society may instill in the average Indian citizen a sense of responsibility about the safety of women around them. Tweets, or Twitter posts, can be used to spread awareness within Indian youth culture and encourage people to take stern action against those who harass women. Tweets typically include photographs, text, and written words and statements that are focused on the safety of women in Indian cities. As a platform for women to express their opinions about how they feel while we go out for work or travel in a public transportation, twitter and other twitter handles that include hash tag messages that are widely shared across the entire globe sir provide women with the opportunity to do so. These women can discuss how they feel when they are surrounded by unknown men and whether or not they feel safe*

***Key Words*: Women Safety, Sexual Assault, Hash Tags, Sentimental Analysis, Tweets on Tweeter.**

1. INTRODUCTION

Several studies have been conducted in cities across India and women report similar types of sexual harassment and passing off comments by other unidentified people. There are some forms of harassment and violence that are very aggressive, including starting and passing comments, and these unacceptable practices are usually seen as a normal part of urban life. According to research that was done in the most populated Indian cities, including Delhi, Mumbai, and Pune, 60% of women report feeling unsafe. Women have the ability to express their thoughts about how they feel when we go for work or ride in a public vehicle thanks to Twitter and other Twitter accounts that feature hash tag messages that are frequently shared throughout the world. These ladies can talk about how they feel and whether they feel comfortable when they are around guys they don't know. Women have the right to the city, which gives them the freedom to go wherever they like, including to places of learning and other places. There are many places in the country where women are still not aware of some of the most basic rights that they can take advantage of in order to empower themselves. This brings us to the next thing that needs the attention of people living in our country. Many women livings in socially and economically backward areas are being victims of domestic violence, without being aware of what they should be doing in order to prevent this from happening and taking a stand for themselves after this happens, women keep on enduring this horrible behavior against them. However, the biggest reason why women feel unsafe in public places like malls is because of girl harassment. They can be preoccupied by their work' attention issues or safety worries as well. Sometimes neighborhood girls would bother the girls as they walked to school, or perhaps there wasn't adequate safety, which would cause young girls to be afraid.

1.1 Motivation

India is experiencing a daily rise in crime. The most concerning sort of crime is crime against women. Women travelling from other nations are likewise in a hesitant state when considering visiting India. Their fear, however, cannot prevent them from participating in any form of social engagement. While there are regulations, there also has to be sufficient safety measures that we must adhere to in order to safeguard women from abuse. A nation cannot advance if women must endure hostility from the populace since women also contribute to the advancement of the country.

1.2 Objective

Strategies, policies, and laws aimed at reducing gender-based violence, including women's fear of crime, are part of women's safety. Safe spaces are necessary for women's safety. Space isn't impartial. Fear-inducing spaces limit movement and the community's utilization of the area.

2. LITRATURE SURVEY

[1] Contextual phrase level polarity analysis using lexical affect scoring and syntactic N-grams:

They offer a classifier that predicts the contextual polarity of subjective clauses in a sentence. We can automatically score the great majority of the words in our input without the requirement for manual labelling thanks to lexical scoring

that was developed from the Dictionary of Affect Language (DAL) and extended through the World Wide Web. To account for the impact of context, they add n-gram analysis to the lexical scoring process. They merged all of the syntactic components from all sentences with the DAL score. Then, as features, extract the n-grams of the sentence's components. The findings indicate a significant improvement over both the easier baseline of lexical n-grams and the baseline for the majority class.

[2] Determining the sentiment of opinions:

Identifying sentiments (the affective parts of opinions) is a challenging problem. They present a system that, given a topic, automatically finds the people who hold opinions about that topic and the sentiment of each opinion. The system contains a module for determining word sentiment and another for combining sentiments within a classifying and combining sentiment at word and sentence levels, with promising results.

[3] Accurate Unlexicalized Parsing:

In this study, we demonstrate that the parsing performance that an unlexicalized PCFG can attain is substantially greater than previously reported, and in fact, far higher than conventional wisdom had considered conceivable. We outline a number of straightforward, linguistically justified annotations that significantly close the gap between a standard PCFG and cutting-edge lexicalized models.

[4] Study of twitter sentiment analysis using machine learning algorithms on python:

People frequently use the social media site Twitter to share their thoughts and emotions on various occasions. Sentiment analysis is a method for analyzing data and locating the sentiments it contains. Twitter sentiment analysis is the use of sentiment analysis to data from tweets on the social media platform in order to derive user sentiments. In this study, we analyze a few studies on sentiment analysis research on twitter, outlining the methodology used, the models used, and outlining a generalized Python-based approach.

[5] Twitter Sentiment Analysis:

Twitter sentiment analysis was created to examine customer perceptions of the essential elements of market success. The application will combine natural language processing methods with a machine-based learning approach that is more accurate for sentiment analysis.

3. METHODOLOGY

This project has been divided into 2 phases.

- First, literature study is conducted, followed by system development. Literature study involves conducting studies on various sentiment analysis techniques and method that currently is used.

- In phase 2 application requirements and Functionalities are defined prior to its development. Also, architecture and interface design of the program and how it will interact are also identified. In developing the twitter sentiment analysis applications, several tools are utilized, such as python shell and notepad.

SENTIMENTAL ANALYSIS

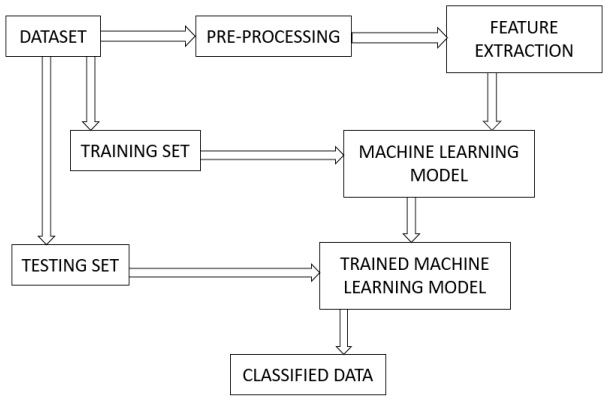

Fig-1: General Methodology for Sentiment Analysis

3.1 SYSTEM ARCHITECTURE

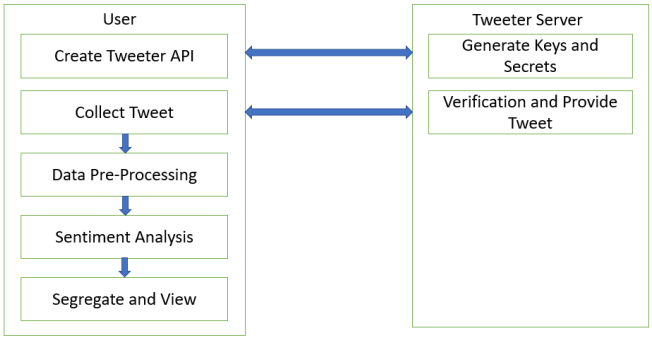

Fig-2: System Architecture

In System Architecture we have user and the tweeter server. Few steps in this have the two-way connections. Like For collecting the tweeter API and collecting the tweet. After creating the tweeter API user sends to the tweeter server where the server generates the keys and send it back. Once the tweets are collected again it sends it to back server where server verifies and again send it back to the user.

3.2 USECASE DIAGRAM

In the use case diagram, we have many connections from user and the tweeter server. Many connections have the 2-way connections. It undergoes many processors like. Requesting tweeter for the generation of the keys. After generating the keys, it verifies the keys.

The user will collect the verified keys and continues the further processing. Like we have data preprocessing, sentimental analysis and segregation.

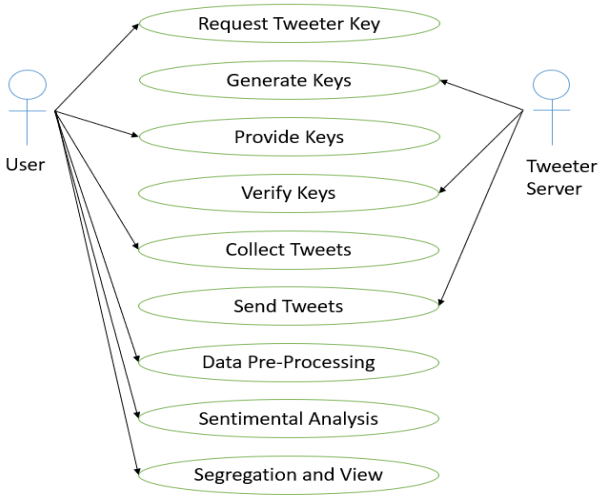

Fig-3: Use case diagram

3.3 SEQUENCE DIAGRAM

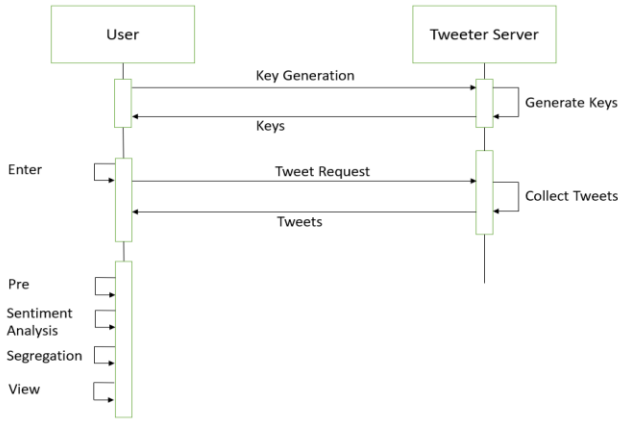

Fig-4: Sequence Diagram

In the sequence diagram we have the same systematic representation as system architecture. It undergoes data preprocessing where all the input tweets will be collected after collecting the tweets will be cleaned and it rescaled dataset through NLP. In the data preprocessing 1 level we will find the sentimental analysis and segregation.

In sentimental analysis the tweets will be classified in to three groups positive, negative and neutral tweets. In segregation classified tweets will be declared for each city based on positive negative and neutral in the form of percentage.

3.4 DATAFLOW DIAGRAM

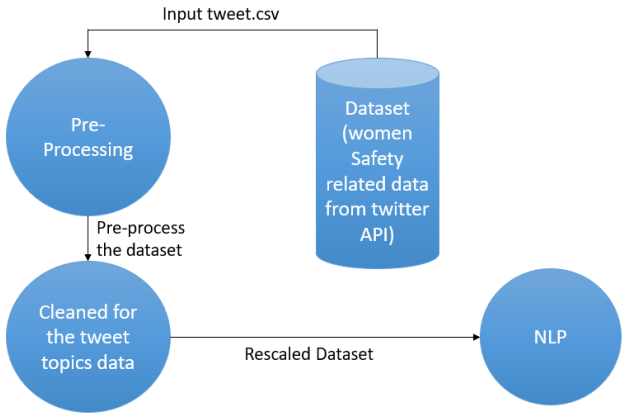

Fig-5: Dataflow Diagram 0

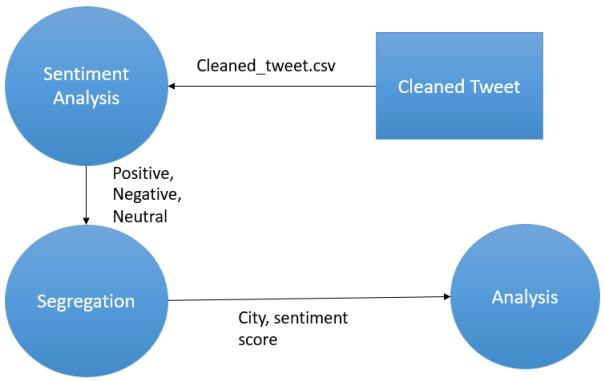

Fig-6: Dataflow Diagram 1

4. IMPLEMENTATION

A subfield of data science known as natural language processing (NLP) involves methodical procedures for intelligently and effectively evaluating, comprehending, and extrapolating information from text data. The large amounts of text data can be organized using NLP and its components to address a wide range of difficulties, including automatic summarization, machine translation, named entity recognition, link extraction, sentiment analysis, speech recognition, and topic segmentation.

- A text is tokenized during the process of tokenization.

- Tokens are words or other items that appear in the text.
- Text objects include sentences, phrases, words, and article

4.1 Text Preprocessing:

As text is the least structured of all the data kinds, it contains a variety of noise and cannot be easily analyzed without pre-processing. Text pre-processing refers to the full procedure of standardizing and cleaning text to remove noise and prepare it for analysis.

It typically consists of three steps:

1. Lexicon normalization.
2. Noise reduction.
3. Object standardization.

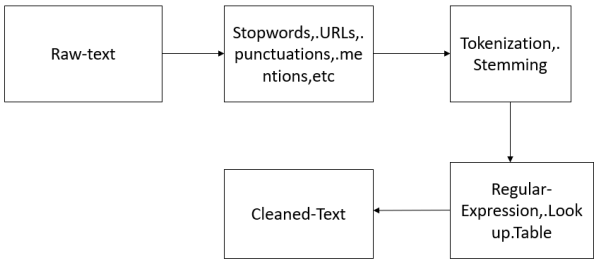

Fig -2: Text Preprocessing

5. PROBLEM STATEMENT

Many incidents of violence and harassment against women and girls have occurred in public locations in different cities, starting with stalking and progressing to sexual harassment or sexual assault. Girls are harassed most often for reasons related to safety or a lack of tangible consequences in their lives. Instead of placing limits on women, society should understand the need of protecting them and that women and girls have the same right to safety in the city as men have.

5.1 EXISTING SYSTEM

On social media, people frequently express themselves openly about how they feel about Indian society and the politicians who assert that women are secure in Indian cities.

People can freely express their opinions on social media networks, and women can publish their stories of sexual harassment they have encountered or how they would have retaliated against it if it had been pushed upon them.

5.2 PROPOSED SYSTEM

Social media can be seen as the ideal medium to discover people's opinions and thoughts regarding various events because people actively communicate and share their opinions on sites like Facebook and Twitter. There are numerous opinion-focused information collection and analytics platforms that try to ascertain people's opinions on various subjects. Twitter posts are brief, and users frequently utilize alternative terms and acronyms. The existing NLP algorithm finds it challenging to easily extract the sentiment from these phrases.

6. CONCLUSIONS

The different machine learning techniques that can help us organize and analyze the enormous amount of Twitter data acquired, including the millions of tweets and text messages posted every day, have been discussed. The SPC method and linear algebraic Factor Model techniques, which help to further categorize the data into meaningful groupings, are two machine learning algorithms that are particularly successful and useful when it comes to evaluating enormous amounts of data. Another machine learning algorithm known as support vector machines is highly popular for extracting useful data from Twitter and gaining insight into the status of women's safety in Indian cities.

7. FUTURE SCOPE

Since only Twitter is taken into consideration in our experiment, we can expand to apply these machine learning algorithms on other social media sites like Facebook and Instagram as well. The proposed ideology can be incorporated into the Twitter application interface to reach a wider audience and perform emotive analysis to millions of tweets to increase safety.

ACKNOWLEDGEMENT

We extend our gratitude to Dr. Thomas P John (Chairman), Dr. Suresh Venugopal P (Principal), Dr. Srinivasa H P (Vice-principal), Ms. Suma R (HOD - CSE Department), Bhavya N J (Associate Professor & Project Coordinator), Ms. Bhavya Balakrishnan (Assistant Professor & Project Guide), Teaching & Non-Teaching Staffs of T. John Institute of Technology, Bengaluru – 560083.

REFERENCES

[1] Agarwal, Apoorv, Fadi Baidsy and Kathleen R. Mckeown. "Contextual phrase -level polarity analysis using lexical affect scoring and syntactic n-grams." proceedings of the 12th European chapter of the association for computational linguistics, associations for computational linguistice,2009.

[2] Barbosa Lucianoand Junla Feng. "Robust sentiment detection on twitter from biased and noisy data." Proceedings of the 23rd international conference on computational linguistics: posters. associations for computational linguistic, 2010.

[3] Bemingham, Adam, and Alan F. Smeaton. "Classifying sentiment in micro blogs: 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 Facebook data: noisy data, large feature vectors, and the role of linguistic analysis:" proceedings of the 20th international conference on computational linguistics association from computational linguistics,2004.

[5] Kim, Soo-min, and Eduard hovy. "Determining the s of options." proceeding of the 20th international conference on computational linguistics Associations from computational linguistics, 2004.

[6] Keindan, and Christopher D. Manning, "Accurate Unlexicalized parsing." proceedings of the 41st annual meeting on association f or computational linguistics-volume 1. Association from computational linguistics,2003.

[7] Charniak, Eugene, and mark Johnson."Coarse-to-fine n-best parsing and maxent discriminative re-ranking". proceedings of the 43rd annual meeting on Associations for computational linguistics. Associations for computational linguistics,2005.

[8] Gupta B., negi M., Vishwakarma., 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 India using twitter data. in computational techniques in information and communication Technologies (ICCDICT),2026 international conference on (pp.525-530). IEEE