Women Protection Analysis Based On Twitter Data Using Ml

Raparthi Shravya[1], Dr.P. Neelakantan[2]

[1]*M.Tech Student, CSE, VNR VJIT, HYD, TS*

[2]*Professor, CSE, VNR VJIT, HYD, TS*

ABSTRACT: Girls and Women have been encountering a ton of savagery and badgering in broad daylight places in different urban communities beginning from following and prompting inappropriate behaviour or rape. This paper examines essential centres around the function of web-based media in advancing the security of ladies in different areas with exceptional reference to the part of online media sites and applications including Twitter stage Facebook and Instagram. This paper additionally focuses around how a feeling of obligation on part of culture can be built up the basic Indian individuals. Tweets on Twitter which typically contains pictures and text and furthermore composed messages and statements which centres around the security of ladies in different urban areas can be utilized to peruse a message among the Youth Culture and instruct individuals to make exacting move and rebuff the individuals who disturb the ladies. Twitter and other Twitter handles which incorporate hash label messages that are generally spread over the entire globe as a stage for ladies to communicate their perspectives about how they feel while we go out for work or travel in a public vehicle and what is the condition of their brain when they are encircled by obscure men and if these ladies have a sense of security? By analyzing the tweets polarity from the Twitter API. In Further improvements, we can use it in any Social Media Platform.

Keywords: Sentiment analysis, machine learning, Bag of words, navy bays

1. INTRODUCTION:

Women Sexual violence occurs throughout the India. Although in most states there has been small research conducted on the problem, available information suggest that in some states around one in four women may experience women sexual violence by an intimate partner [1-3], and up to one-third of adolescent girls report their initial sexual experience as being forced [4–6]. Around 60percent of women's are unsafe while going to office or public places while travelling, mainly in Metropolitan cities of India. Women should feel free and safe while going to working places and public places. Even small girls also faced these harassments while going to tuitions with some of neighbours due to this they are unhappy entire lifetime. So we try to analyse this type of activities to reduce Women Sexual violence.

Social media like twitter, instagram, facebbok are used by various people for their opinions called sentiment. Huge data available in social media sites.Women are also high using Social sites for sharing their incidents or opinions. Sentiment central views of social media content analysis as a result of the most vital data that one desire to extract from the social media content is what users quote and what their opinions' are. These are extracting the core tasks of sentiment analysis. Sentiment analyses are an automatic method of extracting the self-opinionated information.

2. RELATED WORK:

Sentiment Analysis (SA) is a subject of study that investigates people's sentiments, views, assessments, appraisals, attitudes and emotions in the direction of entities such as individuals, services, organizations, issues, products, topics and their characteristics.
It is also known as opinion mining, sentiment mining, subjectivity analysis, review mining, opinion extraction, emotion analysis, etc.

Furthermost of the prevailing approaches have used the terms sentiment analysis and opinion mining interchangeably. According to the research Mathematics, this mining is defined as a quintuple. Sentiment mining = (t, s, h, T)

Where't' is the target opinion, 's' is the sentiment about 't', 'h' is the holder opinion and 'T' is the time. There are 3Approaches in Sentiment Analysis. They are: 1. Machine-learning approach, 2. Lexicon-based approach, 3. Hybrid approach.

Table 2.1 represents the sentimental classification methods.

| Sentiment methods	Classification	Pros and Cons
Lexicon based	Dictionary based. Corpus based. Ensemble approaches.	Pros: Best for domain reliant on, larger-term coverage. Cons: Only Finite number of words in the lexicons
Machine learning Based	Support vector machines. Bayesian networks. Naïve Bayes. Random forest.	Pros: Capacity to adjust and make prepared models for explicit purposes and settings. Cons: Low relevance for new information, since it is important of marked information.
Hybrid based	Lexicon and machine learning based.	Pros: High exactness of new information. Slant vocabulary developed utilizing public assets for assumption discovery. Notion words as highlights in the AI technique. Cons:Noisy data
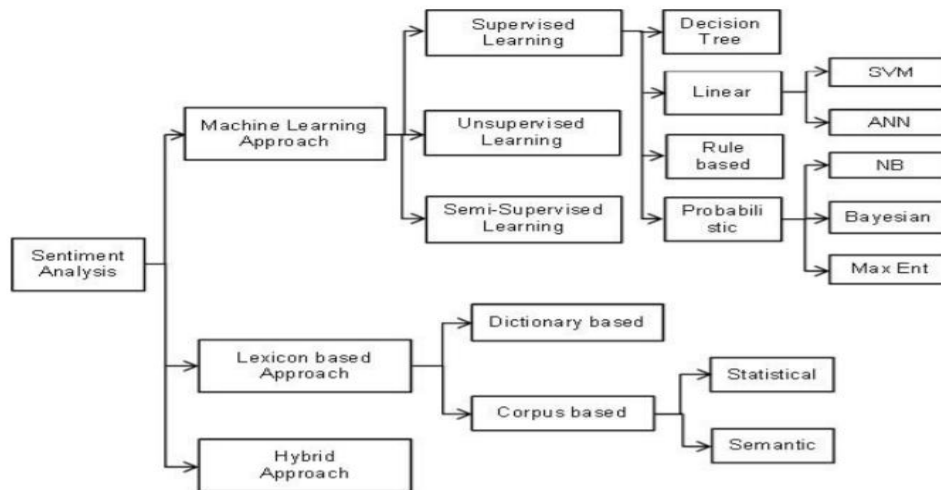
Figure 2.1: Sentimental Analysis Approaches

The follwing table 2.1 describes about different literature that had been done in sentimental analysis with twitter data.

| Authors	Corpus	Method	Comments
Vanzo et al (2014) [7]	Twitter	SVM	Markovian design for context based polarization detection
Shoukry and Rafea (2012) [8]	Twitter	SVM,NB	Sentence level sentiment analyse for Arabic
Duwairi and Qarqaz (2014) [9]	Twitter, FB	Multiple	Rapidminer for text classification on social media reviews
Amolik et al (2016) [10]	Twitter	SVM,NB	movie reviews using machine learning
Kumar et.al 2019[1]	Twitter	SVM	Content based, Bag of words polarity check on 3 cities data.

3. METHODOLOGY:
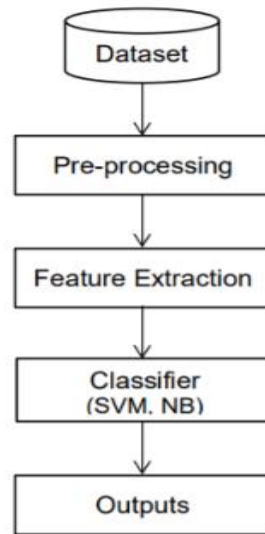
Figure2.2: FlowChart of Machine Learning Algorithm in Sentiment Analysis.

In the Sentiment Analysis the follwing steps are major to identify the positive,negative or neutral of the twitter post.They are:

i.Collecting the Dataset.
ii.Pre-Processing the Dataset.
iii.Feature Extraction.
iv.Apply Classifier.

i.   Data Collection: The data is gathered from the twitter using API. Application program interface (API) is utilized to gather the information. Twitter website is a source which consists of users tweets.

ii.   Data Preparation: In Data Pre-processing removing noisy, unrelated data, inconsistent and incompletedata from the dataset. Generally, in twitter we have to remove URLs, special characters, retweets, hash tags.

iii.   Feature Extraction: In this work, we used Bag of Words to extract features from text documents. After extraction, these features used for training machine learning algorithms. It makes a jargon of the apparent multitude of novel words happening in all the reports in the preparation set. Bag of words features containing term frequencies of each word in each document, i.e. the number of occurrence and not sequence or order of words matters. This can be done by CountVectorizer method in Python

iv.   classification: A classification problem is applied if the output variable is a label or category,   such   as "Rainy" or "Sunny" or "disease" and "no   disease" or   in   our work "Positive" or "Negative".

Algorithm: BagofWords:

```python
from sklearn.feature_extraction.text import CountVectorizer

bow_vectorizer = CountVectorizer(max_df=0.90, min_df=2, max_features=1000, stop_words='english')

# bag-of-words feature matrix
bow = bow_vectorizer.fit_transform(combine['Tidy_Tweets'])

df_bow = pd.DataFrame(bow.todense())

df_bow
```

4. RESULTS:

In this work, we took Twitter API database as input database, after that Pre-process the data set(remove incomplete and noisy data) then apply feature extraction method as bagging of words and finally used Naïve Bayes classification. In this work, we used python language to develop the system. The following figure is the result of the given Twitter API dataset Tweets Accuracy Percentage.
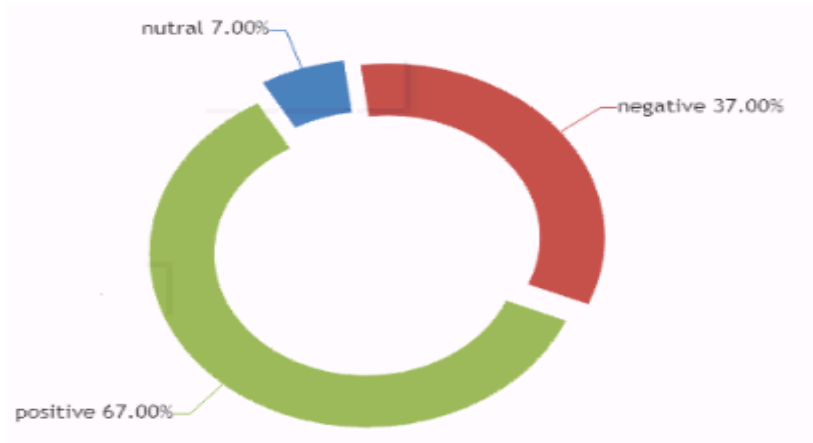

Figure 4:Tweets Accuracy Percentage of Positive, Negative and Neutral.

5. CONCLUSION AND FUTURE WORK:

Throughout this paper we have examined about different Machine Learning techniques that can push us to put together and examine the immense measure of Twitter information. Subsequently we can perform AI calculations to accomplish nostalgic investigation and carry more safety and security to ladies by spreading the mindfulness.

For the future improvement, we can stretch out to apply these Machine Learning techniques on various web-based media stages like face book and integral likewise since in our work just twitter is thought of. Present viewpoint which is proposed can be incorporated with the twitter application interface to arrive at bigger degree and apply wistful examination on a large number of tweets to give more wellbeing.

6. REFERENCES.

[1] Eugene Charniak and additionally Mark Johnson. "Coarse-to-exceptional nbest parsing and MaxEnt discriminative reranking." Process of the 43rd annual assembly on association for computational linguistics. Organization for Computational Grammar, 2005.

[2] Gupta B, Negi M, Vishwakarma K, Rawat G &Badhani P (2017). "Research of Twitter sentiment evaluation making use of system mastering algorithms on Python." International Journal of Computer System Applications, one hundred sixty-five(nine) 0975-8887.

[3] Sahayak V, Shete V & Pathan A (2015). "View evaluation on twitter facts." International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1), 178-183.

[4] Mamgain N, Mehta E, Mittal A & Bhatt G (2016, March). "Belief evaluation of top schools in India making use of Twitter data." In Computational Strategies, in Details and Interaction Technologies (ICCTICT).

[5] Apoorv Agarwal, FadiBiadsy, and additionally Kathleen R. Mckeown. "Contextual phrase-stage polarity analysis using lexical affect racking up and syntactic n-grams." Proceedings of the 12th Seminar of the European Chapter of the Organization for Computational Grammar. Association for Computational Linguistics, 2009.

[6] Luciano Barbosa and additionally Junlan Feng. "Durable view discovery on twitter from biased and loud statistics." Proceedings of the twenty third international convention on computational grammars: posters. Organization for Computational Linguistics, 2010.

[7] Adam Bermingham and additionally Alan F. Smeaton. "Categorizing view in microblogs: is brevity a bonus?." Proceedings of the nineteenth ACM global convention on Details and understanding management. ACM, 2010.

[8] Michael Gamon. "Sentiment category on client feedback statistics: noisy statistics, big function vectors, and additionally the duty of etymological evaluation." Process of the twentieth worldwide assembly on Computational Grammar. Organization for Computational Linguistics, 2004.

[9] Soo-Min Kim and Eduard Hovy. "Identifying the sentiment of reviews." Procedures of the twentieth worldwide convention on Computational Linguistics. Association for Computational Linguistics, 2004.

[10] Dan Klein as well as Christopher D. Manning. "Accurate unlexicalized parsing." Procedures of the forty first Yearly Meeting on Organization for Computational LinguisticsVolume 1. Association for Computational Linguistics, 2003.