Journal of Nonlinear Analysis and Optimization Vol. 15, Issue. 1 : 2024 ISSN : **1906-9685**



UTILISING MACHINE LEARNING TO ANALYSE TWEETS REGARDING WOMEN'S SAFETY

#1 Mrs.R.Vijaya, #2 K.Ramya, #3 K.UMASRIDEVI, #4 E.SIRIVENILA, #5 B.J.V Sai Kiran #6 J.SAIKRISHNA

#1Associative professor in Department of IT, DVR & Dr.HS MIC College of

Technology,Kanchikacherla

#2#3#4#5#6 B.Tech with Specialization of Information Technology , DVR & Dr.HS MIC College of Technology,Kanchikacherla-521180

Abstract In many cities, violence against women, including harassment, is currently prevalent. Stalking is the first step in this, which develops into abusive harassment and assault. In this study, we primarily concentrate on how social media contributes to women's safety in India, paying close attention to the involvement of various social media platforms like Twitter, Facebook, and Instagram. This study also focuses on educating the general public about their obligations to safeguard the safety of women nearby in various Indian cities. A tweet on the Twitter programme can contain text messages, audio data, video data, photos, smiling faces, and hash tags. If there are tweets that are hostile to women, this content can be used to educate people, take strict action, and punish offenders if harassment occurs. Applications that support hashtags, like Twitter and Instagram, enable women to freely express their thoughts and feelings and send messages across the globe. When they commute to work, ride public transportation, or are in a group of shady men, we can find out how they are feeling psychologically and whether they feel safe.

1.INTRODUCTION

Twitter in this modern era has emerged as a microblogging ultimate 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. The data set of the tweet will be used to process the machine learning algorithms and models. This algorithm will perform smoothening the tweet data by eliminating zero values. Using Laplace and porter's theory, a method is developed in order to analyze the tweet data and remove redundant information from the data set. Huge numbers of people have been attracted to social media platform such as Twitter, Facebook, Instagram. People express their sentiments about society, politics, women, etc via the text messages, emoticons and hashtags through such platforms. There are some methods of sentiment that can be classified like machine leaning based and lexicon based learning.

2.LITERATURE SURVEY

2.1 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.

In this research, we offer a method for automatically detecting feelings in Twitter messages (tweets) that takes into account specific features of how tweets are written as well as meta-information about the words that make up these messages. In addition, we use sources of noisy labels as training data. A few sentiment detection websites provided these noisy labels based on twitter data. Our investigations show that because our features can capture a more abstract representation of

JNAO Vol. 15, Issue. 1 : 2024

tweets, our method is more effective than prior ones and also more robust when dealing with skewed and noisy data, which is what these sources deliver.

2.2 Agarwal, Apoorv, Fadi Biadsy, and Kathleen R. Mckeown. "Contextual phraselevel 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

We present a classifier to predict contextual polarity of subjective phrases in a sentence. Our approach features lexical scoring derived from the Dictionary of Affect in Language (DAL) and extended through WordNet, allowing us to automatically score the vast majority of words in our input avoiding the need for manual labeling. We augment lexical scoring with ngram analysis to capture the effect of context. We combine DAL scores with syntactic constituents and then extract ngrams of constituents from all sentences. We also use the polarity of all syntactic constituents within the sentence as features. Our results show significant improvement over a majority class baseline as well as a more difficult baseline consisting of lexical n-grams.

2.3 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.

Microblogging has become a popular method for Internet users to publish thoughts and information in real-time. Automated sentiment analysis of microblog posts is of interest to many, allowing monitoring of public sentiment towards people, products and events, as they happen. The short length of microblog documents means they can be easily published and read on a variety of platforms and modalities. This brevity constraint has led to the use of nonstandard textual artefacts such as emoticons and informal language. The resulting text is often considered "noisy". It is reasonable to assume that the short document length introduces a succinctness to the content. The focused nature of the text and higher density of sentiment-bearing terms may benefit automated sentiment analysis techniques. On the other hand, it may also be that the shorter length and language conventions used mean there is not enough context

3.PROPOSED SYSTEM

In this paper author is describing 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.

In propose work author using TWEEPY 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 this 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.

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.

3.1 IMPLEMENTATION

1) Data extraction: First step involved in analysis of sentiment is the collection of information from the social network website like twitter. This helps in extracting the tweet message but this message also includes extra data like tweets likes, dislikes and comments.

2) Text Cleaning: Once the data is extracted from the twitter source as the datasets, this information has to be passed to the classifier. The classifier cleans the dataset by removing redundant data like stop words, emoticons in order to make sure that non textual content is identified and removed before the analysis.

3) Sentiment Analysis: After the classifier cleans the dataset, the data is ready for the sentimental analysis process. Machine learning

and Lexicon based learning and Hybrid learning are some of the approaches of sentimental analysis. There are also some other approaches such as Nero Linguistic Programming Language and Natural Processing. Training the dataset and then testing that trained dataset involves in machine learning approach. Training data and Testing data are useful for the classifier to perform the algorithm. Maximum Entropy, Naives Bayes classification, Bayesian Networks and Network Support Vector Machine are some of the algorithm which can be used to train the classifier. Testing data is used to identify the efficiency of the sentiment classifier. In case of Lexicon based leaning, training dataset is not used. This approach uses a built-in dictionary in which words associated with sentiments of human are present. The third approach, which is the Hybrid learning, combines both machine approach and lexicon leaning learning approach in order to improve the performance of classifier.

4) Sentiment Classification: At this step, the dataset is ready for the classification. Each and every sentence of the tweet will be examined and opinion will be formed accordingly for subjectivity. Subjective expression sentences are retained and those of objective expression sentences are rejected. Techniques like Unigrams, Negation, Lemmas and so on are used at different levels of sentimental analysis. Sentiments can be distinguished broadly into two groups – Positive and Negative. At this point of sentimental analysis, each of the subjective sentences which will be retained are classified into good, bad or like, dislike or positive and negative.

5) Output Presentation: To generate useful and meaningful information out of the raw data, sentimental analysis plays vital role. Once the algorithm is completed, the outcome of the analysis can be visualized by creating different types of graphs. Bar graphs, Time series and Pie charts are some of the examples which can be used to display the output. To measure the sentiment of the tweets in terms of Positive and Negative, Bar graphs can be used. Similarly, to measure in terms of likes, dislikes, average length of tweet for a certain period, Time series can be used. To obtain the initial source of the tweet, pie charts can be used.

-

3.2 ABOUT DATASET

Helic (sensor - Mar War De	~ 0
e hat yee past house diges how hat hat years ins	
→★・日・日幸臣 米治 告ばる ター・ビー 舟う 田・田・田 + 田 + 田 + 田 和田 日 日 田 白 日 田 田 - 田 田	
- 注意 第2月 100	17.1
Text.	
Bflightsatleft: What he did was manipulative, dishonest, and coercive. Coerced sex ion2v80td0t contentsual sex, it is sexual assault. The bant2vd06	
Pappy 94th Birthday/U/x9flaBelx52 to @_AngelaLansbory Besknolus &: Broomsticks is still one of my bayone movies infutfore's hoping?1x806 https://t.co/C24m/hys/L*	
BRestonTishen: Two years ago #MeToo woke up our nation. It's become part of our national consciousness, demanding better for survivors24806*	
BHING15: #Timestip #WeTooVoter #WenToo #Wetoo #Elexit #WednesdayWisdom InTime has come that laws should be amended so that manpant take2v806'	
Bambertandayn: Months behrer Me Top broks is wide open, a man tapped me on the shoulder at a ressament and said he had read this op-ed (2x000 Bmarny: otawa: What a year on twitter: (2x80x90ve seenintoLiberatini). Lawyers: defend attack on Rule of Law. PObstoctionOfJastice Init2. Female/2x806	
(2) Entering of participation with a year on lawer indexestive intermination. Lawyork and the state of the	
Child Andrease weiges in on New Conton across sex Address investigation - Year, Connection high style by Article across in high style across and across and across and across and across across and across ac	
procession with the method of the second sec	
CourseManager white maker in secondly ponder, cuture or creat. Help Ouser Ministry upit your community in low, the believe lacked01 CourseManager white maker is secondly ponder, cuture or creat. Help Ouser Ministry upit your community in low, the believe lacked01	
THETOO and NE TOO https://t.co/rRCnCH23W	
Bugarukeresience, BillacheintheOC I'm not surprised either. Every so often one of hese threads pape up like a mini-testpa moment, and gen/24006"	
(Pha2012: Left wing success is a shit hole to the rest of us #MpTop #McTopMovement #samish #MACA #Bexa #Wa&Away #WeBulliTheWall #KAC2/b806"	
Binarry, otawa. What a year on milter- Dis80x99/ve seenvicil. Lewyers - defend attack on Rule of Law - POtstructionDL/Lustee Intr2. Feminis2/a806	1
Berrygards: Can we PLEASE more away from focusing so much on health care in the next debate. We all know where they stand by now (mye wan7)/e806"	
PreneWoodbury, Jen &, Coton + Brandi &, Matt + Zane = Non-Stop Drama I Pop-Cut Girl, Al #TateredCover Book Stores, Derver1/r#GreatReads2/aB06	
Bidavelgamithphd: Men's current AMel'op Systema Ses men living in fear at work. Curing men's Gyrophobia -abnomial lear of women-requires/2x8001"	
(RealCandaceO: I think Matt Lauer is a scorebag/2x80/x94 as a father and husband. Bus I also think his accuser/2x80/x94a 30 yr old woman who clearly exch/2x8067	
dme1ster @kaceytrum @ladygaga @Ninja Then cancel culture and metpo, he played this right. Ahead of the curve."	
Relightick in 1961 ??? Really??? So because of something he said 40 years ago you?ix80x90re going to chantise him? You?ix80x90re?u106 https://t.co.WhotLeBv?U	
BLisaEntion: During this time of McTop. It is also very very very very very very very very	
Broanty stlaws What a year on twitter Utst/fa/Hwy secritical.jberal/in1. Lawyers - defend attack on Fule of Law - PObstructionOfJustice Init2. Femine24000	
@McDonalds, it is your responsibility to take real steps to ensure safe workplaces. You2/a80x/6/ere making our need for uni2x806 https://cpi/2GbPvcpdz/	
@RituG15: #TimesUp MeTopYoter #MenTop Metopy #Brent #WednesdayWedom InTime has come that taxes should be amended so that rampant take21x806"	
Branny otawa What a year on twitter 12bd/0x99ve seeminial benahrd. Lawyers - defend attack on Rule of Law PObatractionOUssice Initia 2x0000	
b"@RachelintheOC Tm not surprised either. Every so often one of these threads pops up like a mini-metoo moment, and 2x806 https://t.co/BPta/CaEa//"	
@davidgamidgabd: Men's current AMeTop hysteria has men living in loar at work. Curing men's Cycoghobia-abromual lear of women-requires2x8000"	
And what inspect all that owners still wanted you'nto love them bila Carol Diehi, V#MeToo #CarolDiehi	
glussBritton: During this time of Me Loo. It is also very very very very very very very very	
@manny_otawa: What a your on twitter: IDx80/v90ve seeninkLiberal/in1. Lawyers - defend attack on Rule of Law PChatactionDQustee hin2. FeminisZv80vF	
Be almost the Lemma Lemma for a last of the ARE Tax backback is a matched last to the Mark is bad for the ARE Tax backback is a matched last to the ARE to bad for the ARE matched and base To ARE	1.473
133 🔶 Mila Jaan	

Fig 4.1 Dataset information

5.RESULTS AND DISCUSSION Analysis of Women Safety in Indian Cities Using Machine Lasming on Tweets



Fig 5.1 In above screen click on 'Upload Tweets Dataset' button and upload tweets

		8		
🕆 🛄 🗤 WomenSafteyMachineLear = datasat	v O Saisth dealed	P	n Cities Using Machine Learning on Tweets	
a - Nov-Addar		O		
StockMarket/ ** Newye	Option would lived	7,94		
StatestPerio	21-86-2500 (96:28	Text Document .		
TeMRockMe Q MeTrio Investigan	19-19-2019 16-21	Microsoft Ivent [Women Saftey Graph	
TiefficSige		_		
Tratfleputati VESHARE				
Webswee WdmCempn				
Viticulture				
Web4ttacks				
Worrentlefte				
(staset				
Elemente MeTeo, Iweets ov	10			
	Open	Catof		

1315

Fig 5.2 In above screen uploading MeeToo_tweets.csv file and then click on 'Open' button to load dataset and to get below screen

Read Tweets Insathirth: "What he 94th Birthday@a90: LoriTrahan: Twe ye		and the second se	ng Algorithm Women Saftey Graph	
94th Birthday @a90:		a second design of the second s		
bortamblyu: Mourto I nuy_ottawa: What a y Advocate Weighs in o ghranthirth: "What he ialiststeved: I would i or and ME TOO http: 'oo and ME TOO http:	hefore MeTos brok year on twitter: 121x2 on New London Scho did was manipulativ instead tell my 2 dan r queer in secondity, ps:Pt.co/yRCaGHrM	e as wide open, a man tapped me on the Wat9re scenario Liberal'mi. Lawyers, ed Suz Abuse Investigation NBC Cos- «, diskoaset, and coercive. Coerced to ghtere winDuring the time of the # genter, winDuring the time of the # A'	on 'nTime has come that laws should be amended to that rangpart fake2/x306' e shoulder at a restaurant and and he had reach this op-of 12/x306' defend attack on Rule of Law. #ObstractionOfDustice win2, Fensinis2/x306' meetical https://t.co/YyX7#FeciO2/x306 https://t.co/YXUL7EVxDP' xim2x50/x99t consensual wey, it is seemal avaant. The hard2/x306' MeTon norvenout it is very very very very very 2005/2006' Ministry uplift your community is love. We believe fai2/x306' these threads page up like a mini-metoe norment, and gen2/x306'	

Fig 5.3 In above screen each line represents one tweet and you can scroll down above screen text area to view all tweets. In above screen we can see all tweets contains special symbols and stop words and to clean those tweets click on 'Tweets Cleaning' button

		Analysis of Women Sa	fety in Indian C	ities Using M	lachine Lei	truing on Tv	veets		
Upload Tweets D	ataset	E#krest/WomenSafteyM	achineLearning	dataset/McTo	o_tweets.cs				
Read Tweets	Tweets Cleaning	Run Machine Learni	ng Algorithm	Women Sa	ftey Graph				
ightsathirth manipulative d Predicted Sentiment : NEG. Polarity Score : 0,1		es consensual sex sexual assa	alt						
appy angelalanthury bodk redicted Seatiment : POSI 'slarity Score : 0.65		ne favorite noviesnikeres							
ploritrahan two years ago redicted Seatiment : NEUT slarity Score : 0.5		part autienal consciousness d	enanding better						
nemp metoevoter mentoe redicted Sentiment : NEG idarity Score : 0.0		sdom ntime come laws amende	d rampant						
nhertaublyn noutle neto rodicted Sentiment : NEG. slarity Score : -0.05		apped skoulder restauraut said	l read oped						
annyottawa year twitter la redicted Sentiment : NEG, ularity Score 20,0	wyors defend attack rule la ATTVE	v obstructionoljusticv							
hild advocate weight new h	ondon uchool sex abuve invo	estigation abe connecticut							
O Type here to sea	in g	0 2 🕯 💼					£. 0	1421	P

Fig 5.4 In above screen each tweet having tweet text and then displaying tweets sentiments with polarity score. Scroll down above text area to see all tweets. Now click on 'Women Saftey Graph' button to get below results and by seeing that result user can easily understand whether area is safe or not. If area is safe then more peoples will express either positive or neutral tweets and if not safe then more peoples will discuss negative tweets.

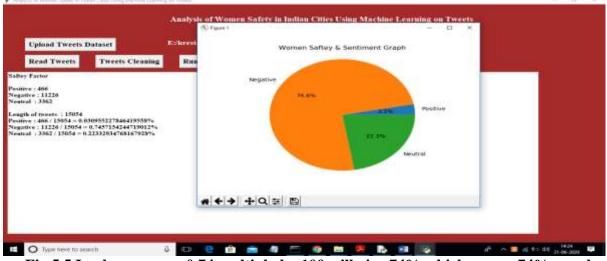


Fig 5.5 In above screen 0.74 multiply by 100 will give 74% which means 74% peoples are talking negative and area is not safe and only 22 and 3% peoples are talking positive and neutral.

6.CONCLUSION

Throughout the research paper, we discussed a number of machine learning techniques to assist us in organising and analysing the enormous amount of Twitter data we have gathered, which includes millions of tweets and text messages posted every day. When it comes to analysing enormous amounts of data, the SPC method and linear algebraic Factor Model techniques, which aid in further categorising the data into meaningful groupings, are two machine learning algorithms that are particularly effective and helpful. In order to understand the state of women's safety in Indian cities, support vector machines, a type of machine learning technique, are frequently used to extract useful information from Twitter..

1317

REFERENCES

[1]. Apoorva Agarwal, Fadi Biadsy, and Kathleen R. Mckeown. "Contextual phraselevel 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]. Luciano Barbosa 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]. Adam Bermingham 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]. Michael Gamon. "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]. Soo-Min Kim and Eduard Hovy. "Determining the sentiment of opinions." Proceedings of the 20th international

conference on Computational Linguistics. Association for Computational Linguistics, 2004.

[6]. Dan Klein and Christopher D. Manning. "Accurate unlexicalized parsing." Proceedings of the 41st Annual

Meeting on ';'Association for Computational Linguistics

[7]. Eugene Charniak 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.

JNAO Vol. 15, Issue. 1: 2024

[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.

Author's Profile

#1:-Mrs.R.Vijaya working as Associate Professor in Department Of AI & IT in DVR & Dr.HS MIC College Of Technology,Kanchikacherla-521180.

#2:-K.Ramya(**20H71A1228**)B.Tech with Specialization of Information Technology in DVR & Dr.HS MIC College Of Technology,Kanchikacherla-521180.

#3:K.UmaSridevi(20H71A1255)B.Tech with Specialization of Information Technology in DVR & Dr.HS MIC College Of Technology,Kanchikacherla-521180.

#4 :- E.SiriVenila(20H71A1241) B.Tech with Specialization of Information Technology in DVR & Dr.HS MIC College Of Technology,Kanchikacherla-521180.

#5:-B.J.V.SaiKiran(20H71A1212) B.Tech with Specialization of Information Technology in DVR & Dr.HS MIC College Of Technology,Kanchikacherla-521180.

#6:-J.SaiKrishna(20H71A1234) B.Tech with Specialization of Information Technology in DVR & Dr.HS MIC College Of Technology,Kanchikacherla-521180.