IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A REVIEW ON SENTIMENT ANALYSIS: SOCIAL MEDIA ABUSE TO WOMEN

[1]Syed Zebanaaz, [2]Dr. Mukta Dhopeshwarkar

[1]Research Student, [2]Assistant Professor

[1]Department of Computer Science & Information Technology,

[1]Dr. Babasaheb Ambedkar Marathwada University, Aurangabad(Maharashtra), India

Abstract: Internet being the basic necessity for humans, plays a significant role in all walks of life as it has become a platform for online learning, exchanging ideas and sharing opinions. The surge in social networking applications has allowed people to use these platforms to voice their opinions for daily issues. This has resulted into in depth work in the field of sentiment analysis of twitter data. Sentiment analysis also referred to as opinion mining is a general form of dialogue preparation task that aims at discovering the sentiments behind opinions in texts on varying topics. Many recent studies in the field of sentiment analysis have also diverted the mind of researches towards daily societal issues. The Social Networking application Twitter is considered to be a popular microblog. This platform allows users to comment or tweet their opinions on a topic of their choice. Much attention is being given to Research on Opinion Investigation or Sentiment Analysis of Twitter data over the last decade which involves dissecting "tweets" (comments) and the content of these expressions by multiusers on varied topics. This review paper emphasises on how women are abused on social networking platforms. The levels of sentiment analysis and approaches of sentiment analysis are also discussed.

Keywords **- Sentiment analysis, opinion mining, lexicons, machine learning, hate speech, vulgarity.**

I. INTRODUCTION

The recent era of Internet has changed the way how people express their views or share their opinions. The use of social network sites like Facebook, Twitter, Google Plus, etc. have popularized so much as millions of people are using these platforms to express their emotions, opinions and share views about their daily activities. Social media is generating a large volume of sentiment rich data in the form of tweets, status updates, blog posts, comments, reviews, etc.

The Sentiment analysis technologies facilitate the automatic analysis of the information distributed through social media to identify the polarity of posted opinions [1]. These technologies have been widely used in the last years to analyze other aspects, such as the stance of a user towards a topic [2] or the users' emotions [3], even combining text analytics with other inputs, including multimedia analysis [4] or social network analysis [5].

Although there are number of existing social networks, Twitter ranks as one of the leading platforms and tends to be one of the most important data sources for researchers. Twitter is a well-known real-time public microblogging network where, frequently, news appear before than on official news media.[6] Characterized by its short message limit i.e. 280 characters and unfiltered feed, its being widely used, especially amid events.

Language though reveals the values of people and their perspectives or mindset. As the current trend of Social networking usage, Twitter is also used to spread hate messages. Hate speech can be referred to as a kind of speech that unfairly criticise a person or multiple persons. This criticism is usually defined by race, ethnicity, sexual orientation, gender identity, disability, religion, political affiliation, or views. The Rabat Plan of Action of the United Nations defines the guidelines to differentiate between free speech and hate speech, and also recommends differentiating between three types of expressions: i) expression that constitutes a criminal offence; ii) expression that is not criminally punishable, but may justify a civil suit or administrative sanctions; iii) expression that does not give rise to criminal, civil or administrative sanctions, but still raises concern in terms of tolerance, civility and respect for the rights of others." [6]

Online hate also described as abusive language, aggression, cyberbullying, hatefulness, insults, personal attacks, provocation, racism, sexism, threats, or toxicity, has been identified as a major threat on online social media platforms.[7]

Undoubtedly, the use of social media, might expose people to threats such as cyberbullying, which is considered to be one of the most significant social attacks happening on social media platforms these days. Cyberbully is defined as posting offensive messages against an individual through digital means, often anonymously [8]. The negative consequences of cyberbullying are not to be ignored, as it may affect the victims' mental health, causing certain psychological conditions such as depression, low self-esteem, disability, social anxiety, suicide, fear, poor social relations with peers, and some studies also indicate the emergence of cases and complaints of pain, headache, stomach pain, difficulty in sleeping and other physical symptoms [9]. Thus, the need for detecting hate speech on social media is an important task for individuals and society.

II. LEVELS AND APPROACHES OF SENTIMENT ANALYSIS

2.1 Levels of Sentiment Analysis:

It is important to highlight the fact that sentiment mining can be performed on three levels as follows:

2.1.1. Document-level sentiment classification:

This level is used to classify a complete document as "positive", "negative", or "neutral". And thus texts which comprise comparative learning cannot be considered under this level.

2.1.2. Sentence-level sentiment classification:

This level classifies each sentence as "positive", "negative" or "neutral". If a sentence depicts no opinion it falls under neutral category. This level of analysis relates to subjectivity classification. The subjective statement displays the polarity of an entity either in affirmation or negation i.e. good-bad.

2.1.3. Aspect and feature level sentiment classification:

Aspect level is used for a detailed analysis. The core task of this level is to identify aspect of the text. At this level sentiment analysis becomes two level task i.e. finding the aspects in the text and then classifying them into respective aspects.

2.2. Approaches:

Sentiment Analysis techniques can be broadly classified into Lexicon based approach, Machine Learning approach and hybrid approach.

2.2.1. The Machine Learning Approach (ML):

Implements the widely used Machine Learning algorithms and it uses linguistic features.

2.2.2. The Lexicon-based Approach:

It is dependant on a sentiment lexicon. The term Lexicon can be defined as a collection of known and precompiled sentiment terms. It can be further divided into a dictionary-based approach and corpus-based approach. These approaches make use of semantic or statistical methods to find out the sentiment polarity of the text.

2.2.3. The Hybrid Approach:

As the name suggests is a combination of both the approaches and the sentiment lexicons play a key role in the majority of methods.

III. LITERATURE REVIEW

The violations of human rights are noted to be perpetrated against men and women [10]. The consequences of these violations, however, vary according to the victim's gender.

Studies mentioned in [11] have shown that the aggressive acts carried out on women differ, exhibiting certain traits.

Raisi [12] has proposed a model that detects offensive comments on social networks, in order to intervene by filtering or advising those involved. To train this model, comments with offensive words from Twitter and Ask.fm were used.

An architecture by Chen [13] was proposed to detect offensive content and identify potential offensive users in social media. The system achieves an accuracy of 98.24% in sentence offensive detection and an accuracy of 77.9% in user offensiveness detection.

Researches in [14] used Bag-of-Words (BoW) in offensiveness detection. The BoW approach treats a text as an unordered collection of words and disregards the syntactic and semantic information.

N-gram approach is considered to be one of the improved approaches as in that it brings words' nearby context information into consideration to detect offensive contents [15]. N-grams represent subsequences of N continuous words in texts. Bi-gram and Tri-gram are the most popular N-grams used in text mining. As user-level detection is a more challenging task and studies associated with the user level of analysis are largely missing.

Pendar [15] in his research has used lexical features with machine learning classifiers to differentiate victims from predators in online chatting environment.

A rule-based communication model to track and categorize online predators have been proposed by Kontostathis et al [16].

Pazienza and Tudorache [17] propose utilizing user profiling features to detect aggressive discussions. They use users' online behavior histories (e.g., presence and conversations) to predict whether users' future posts will be offensive or not.

The first of the two widely used datasets was the research work of Waseem and Hovy [18]. In particular, the dataset is comprised of 16,914 annotated tweets, of which 3383 are categorized as "sexist" and 1972 as "racist".

The second dataset, comprised of 6909 annotated tweets, was introduced in Waseem [19]. This dataset includes 3000 tweets from the previous dataset, albeit with new annotations, and 4000 new tweets.

Badjatiya et al. [20] used the 16,000 tweet dataset which was provided by [18].

IV. CONCLUSION

Sentiment analysis plays a huge role to understand people perception and helps in decision making. The systematic literature review put forth provides information on studies done on sentiment analysis in social media. Online hate is an uncontrolled problem faced these days, with the negative consequence by prohibiting an individuals participation in online discussions and causing cognitive harm to them. The present research provides a novel exploration of women experiencing sexual harassment by strangers or known personnel's in an online context. This review focuses on how women generally find the experience of online sexual harassment to be unwanted and unpleasant, with many downstream negative consequences. The most popular social media site to extract information is Twitter. Most of the papers reviewed use twitter as their social media platform. This is due to the availability, accessibility and richness of Twitter content as we can get millions of tweets every day on almost any topic.

V. FUTURE SCOPE

This is a preliminary work to show how sentiment analysis can be helpful in predicting online hate speech that women face. As future work we will try to collect the tweets and create a dataset from twitter and then analyse which tweet is abusive for women and the comment made is in between friends or by any male with intent to harass a woman.

VI. ACKNOWLEDGEMENT

I am thankful to CS and IT Department of Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) for providing me a moral support.

REFERENCES

[1] Liu, B. 2012. Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol., 5: 1–167.
[2] Mohammad, S., Bravo-Marquez, F., Salameh, M., Kiritchenko, S. Semeval. 2018. Affect in tweets. In Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, LA, USA, 1–17.
[3] Cambria, E., Poria, S., Hussain, A., Liu, B. 2019. Computational Intelligence for Affective Computing and Sentiment Analysis [Guest Editorial]. IEEE Comput. Intell. Mag. 14: 16–17.
[4] Li, Z.; Fan, Y., Jiang, B., Lei, T., Liu, W. 2019. A survey on sentiment analysis and opinion mining for social multimedia. Multimed. Tools Appl. 78: 6939–6967.
[5] Sánchez-Rada, J.F.; Iglesias, C.A. 2019. Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. Inf. Fusion. 52: 344–356.
[6] Juan Carlos Pereira-Kohatsu, Lara Quijano-Sánchez, Federico Liberatore, Miguel Camacho-Collados. 2019. Detecting and Monitoring Hate Speech in Twitter. Sensors: MDPI.
[7] Joni Salminen, Maximilian Hopf, Shammur A. Chowdhury, Soon-gyo Jung, Hind Almerekhi and Bernard J. Jansen. 2020. Developing an online hate classifier for multiple social media platforms. Human Centric Computing and Information Sciences.
[8] Merriam-Webster. 2019. Social Media | Definition of Social Media by Merriam-Webster.
[9] Smith, P.K., In Jimerson, S.R., Nickerson, A.B., Mayer, M.J., Furlong, M.J. 2011. Cyberbullying and Cyber aggression. (eds) Handbook of School Violence and School Safety: International Research and Practice. Routledge, New York.
[10] M. R. Decker, A.-L. Crago, S. K. Chu, S. G. Sherman, M. S. Seshu, K. Buthelezi, M. Dhaliwal, and C. Beyrer. 2015. Human rights violations against sex workers: burden and effect on hiv, The Lancet, 385(9963): 186–199.
[11] S. A. Basow, K. F. Cahill, J. E. Phelan, K. Longshore, and A. McGillicuddy-DeLisi, 2007. Perceptions of relational and physical aggression among college students: Effects of gender of perpetrator, target, and perceiver. Psychology of Women Quarterly, 31(1):85–95.
[12] Raisi, E., Huang, B. 2016. Cyberbullying identification using participant-vocabulary consistency. arXiv. arXiv:1606.08084.
[13] Chen, Y., Zhou, Y., Zhu, S., Xu, H. 2012. Detecting offensive language in social media to protect adolescent online safety. In Proceedings of the 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conferece on Social Computing, Amsterdam, The Netherlands. 71–80.
[14] A. McEnery, J. Baker, and A. Hardie. 2000. Swearing and abuse in modern British English. In Practical Applications of Language Corpora Peter Lang, Hamburg. 37-48.
[15] N. Pendar, 2007. Toward spotting the pedophile telling victim from predator in text chats. In Proceedings of the First IEEE International Conference on Semantic Computing. 235-241.
[16] A. Kontostathis, L. Edwards, and A. Leatherman. 2009. Chatcoder: Toward the tracking and categorization of internet predators. In Proceedings of Text Mining Workshop 2009 held in conjunction with the Ninth SIAM International Conference on Data Mining.
[17] M. Pazienza and A. Tudorache. 2011. Interdisciplinary contributions to flame modeling,. AI* IA 2011: Artificial Intelligence Around Man and Beyond. 213-224.
[18] Waseem, Z., Hovy, D. 2016. Hateful symbols or hateful people? Predictive features for hate speech detection on twitter. In Proceedings of the NAACL Student Research Workshop, San Diego, CA, USA. 88–93.
[19] Waseem, Z. 2016. Are you a racist or am i seeing things? annotator influence on hate speech detection on twitter. In Proceedings of the First Workshop on NLP and Computational Social Science, Austin, TX, USA. 138–142.
[20] Badjatiya, P., Gupta, S., Gupta, M., Varma, V. 2017. Deep learning for hate speech detection in tweets. In Proceedings of the 26th International Conference on WorldWideWeb Companion, International WorldWideWeb Conferences Steering Committee, Perth, Australia. 759–760.