Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 7, July : 2024

SURVEY ON IDENTIFICATION OF WOMEN SEXISM AND CRITICISM IN SOCIAL MEDIA USING MACHINE LEARNING TOOLS.

Praveena Pillala[1]*, Kondapalli Venkata Ramana[2]
[1] Department of CSE-Centurion University of Technology and Management, Visakhapatnam, India; Email: praveena.aucoe@gmail.com
[2] Department of CS&SE-Andhra University College of Engineering, Andhra University, Visakhapatnam, India; Email: kvramana.auce@gmail.com

Abstract:

In the present state women's sexism and criticism are highly depicted in the social network. The women are commented on through social media mostly on their physical appearance, sexism, criticism, etc. Women often face criticism and harassment on social media that is rooted in sexism and misogyny. It can also take many forms, including verbal abuse, physical shaming, sexual harassment, and discrimination in the workplace or other areas of life. It is necessary for individuals and social media platforms to recognize and analyze to prevent impoliteness towards women fraternity. Women are much scarred about sexism and criticisms are orderly increasing day by day in online environments and online community groups. A few present research studies identified more such cases on various social media platforms like; Face book, Twitter, Instagram, Snapchat, Pinterest, Reddit, etc which is an important matter to be concerned. This study is related to the most published data on women's sexism and criticism are being identified using advanced automation tools like natural language processing and machine learning techniques and are analyzed with clear extraction from the social media data resources to detect the impoliteness.

Keywords: Sexism and Criticism, Social Media Platforms, Sentiment Analysis, Natural Language Processing, ML Classification.

Introduction:

In the present day, women are not well represented in social media networks. The present phase of women's participation in any form of development and socialistic activities is much scared in social media platforms. The voice of women on social media platforms going down drastically due to lack of traditional values, cultural norms, human values, laws, economic systems, and insecure social media presentations. The women are distressed through social media and blamed by different societal groups. The major cause of the phenomenon due to women's sexism and criticism is the suppression of women on social media platforms. It is time to protect and secure women's rights in social media. It is important for individuals and social media platforms to recognize and address the needful rights to protect women's situation on a stand and reinforce the idea that women need respect, protection, and compliments.

Figure: 1 Women in Social media platforms.

Figure 1 depicts the fear and anxiousness of women barbarically surrounded by social media platforms. The women are social media platform where women mainly refer to gender-based discrimination that is unbreakable in societal structures. Sexism is characterized by clearly negative attitudes and misbehaviors towards women, such as verbal criticism, abused words, belittling or demeaning and many more threats [1]. Women face a lot of criticism on social media platforms, which can range from personal attacks on their appearance and manners by wrongly questioning their opinions and ideas. This criticism is often rooted in sexist attitudes and bias, which affects stereotypes and discrimination against women. Women who speak out on controversial topics or the real challengers who can establish the norms and beliefs which are often subjected to a higher level of scrutiny and criticism counterparts above the men. So this can make it difficult for women to express their opinions and participate in online discussions. Considering these input features presents research in reviewing the problem statement with recommended advanced tools allied with Sentiment Analysis in Natural Language Processing, Machine learning techniques and BERT are proposed for the further research. The Identification and Classification of these two problems are evaluated with various techniques, tools, datasets, features, and output summarization studies.

The main contributions of this work are,

- Studied technique-wise, dataset-wise, platform-wise related works on women sexism and criticism conducted by researchers of different countries.
- Studied the outcome of different machine learning, deep learning algorithms and analysis of their performance.
- Proposed an esteemed model for better identification of women sexism and criticism on social media platforms through proper preprocessing technique.

Fatimah Alkomah et.al [2] the paper states the enhanced review on main datasets, textual features, and machine learning models of textual hate speech detection systems are highlighted in this work, which also conducts a thorough review of these systems. Machine Learning Hate Speech Models is used to implement through TFIDF Methods, Lexicon-Based Methods, Deep Learning Methods, and LSTM models, BiLSTM hybrid models, are used to analysis the hate speech identification. BERT is a transformer-based machine learning approach for natural language processing that Google based on information derived from plain text. The BERT Model technique covers around 33% of study, whereas LSTM and CNN each cover about 20%. The findings of this literature assessment and content analysis are combined to produce numerous themes for 138 relevant pieces. The machine learning model's architecture and the lack of consensus on the definition of hate speech are two of the main challenges to hate speech detection.

Md Saroar Jahan et.al [3] the study presents assessment of the literature with an emphasis on deep learning and natural language processing technologies, key techniques highlighting the terminology, and processing pipeline used with a focal point on deep learning architecture. Based on methodological standpoint PRISMA guidelines follow the systematic reviews of literature collected from the ACM Digital Library and Google Scholar. In order to provide the researcher an extensive and worldwide approach in the field of automatic hate text (HT) detection, the variety of technology, applied domain, and context-related factors require regular up-to-date of latest developments in this field. Similarly, deep learning techniques like BERT, LSTM, and CNN, where BERT's rising importance is highlighted more.

Siti Ummi Masruroh et.al [4] has stated that, the National Commission on women's safety from Indonesia has taken into the consideration on violence against women in Twitter and other social media networks. This study has examined the safety of women on Twitter, regarding the government of Indonesia raising the Bill on the Elimination of Sexual Violence in Indonesia's Twitter platform. A separate group is appointed to see the public tweets and verify whether the text is positive, negative, or neutral. By applying Natural Language Processing (NLP) using sentimental analysis the sexual violence tweets are identified. Using the algorithms like Support Vector Machine (SVM) with 97%, and Naive Bayes Classifier (NBC) with 94.50% the data are classified with positive and negative classes.

Elisabetta Fersini et.al [5] studied the frequent exponential growth of sexism in social media as a form of discrimination against women on the internet. Memes are often formed of text and visual elements which can hurt women's psychology. Particularly in the form of memes which create a lot of disturbance in social media. Memes can carry messages like ranging, objectification, violence, and shaming which can typecast the women. In the year 2017, study conducted by Amnesty International found that 46% of women had suffered misogynistic or sexist online harassment on social media. About 4000 women involved between the ages of 18 to 55 were interviewed regarding cyber sexism in 8 different countries. The main characteristics of meme statistics are identified through the uni-model and multi-model classifiers approach which are used for labeling both text and visual images. Finally, results are presented in terms of both micro and macro performance of both textual and visual representation.

D. Madhubala et.al [6] stated that, women in well-known positions including athletes, politicians, celebrities, and common women are receiving threats of violence on social media. There are three different types of sentimental analysis for the separation of positive, negative, and neutral labels that are classified using machine learning algorithms. The hashtag is used to extract data from the user using keywords like the hashtag and emoji-based text classification. The sentiment classification of the user post also includes the hashtag. The OMSA (opinion mining and sentiment analysis) is examined in terms of the technical side employing techniques like key-based classification, and lexicon-based classification for the result of the sentiment analysis. According to the testing results using the SVM method, accuracy was 90.7%, which is an estimable performance.

Francisco Rodriguez-Sanchez et.al [7] this paper study states sexist behavior can notice in many various forms, including hidden stereotypes assumptions, and attitudes that are immensely destructive to women and society even when they go unnoticed. The first dataset of sexist tweets and attitudes from Twitter in Spanish dataset is used for the investigation. In this study, we investigate the ways in which sexist attitudes, behaviors, and beliefs are communicated via Twitter. It is mostly based on supervised machine learning on various textual properties such as unigrams and bigrams, sentiment-based data, or syntactic categories. For each keyword expression, we only collected up to 15,000 tweets. Only two classes "misogynous" and "not misogynous" are taken into account in the 3307 tweets that make up the AMI dataset used for this experiment (training set). According to the experimental findings, BERT beats the other algorithms evaluated, detecting sexist expressions with an accuracy of 74%.

Gregorio Arturo Reyes Gonzalez et.al [8] this paper study states using supervised algorithms, data scientists have focused primarily on cyber bullying and language pattern recognition. Sentiment analysis is thought to be able to connect mindset to certain topics, which will help find instances of possible digital violence against women. The two basic categories of sentences are objective sentences which contain information about facts, and subjective sentences which convey expressed ideas, attitudes, approaches, and viewpoints. To analyze affective states and subjective data as part of sentiment analysis, we use natural language processing (NLP), text, and computational linguistics. The Hateval training dataset from GitHub is used as the training model in the research on sentiment analysis that uses the

Twitter dataset as a data source. The results that are represented in 2D visualization have shown that Digital violence works against women on social media platforms.

Research Design:

The research design identified the gap in which the scope of the researcher is finding the negative meaning of sentences on women's activity on social media platforms. The initial process begins with data collection and identification which contains the written text or transcripts of verbal communication that criticize the women [9]. Social media has become an integral part of modern life, connecting people across the globe and facilitating the exchange of ideas, information, and emotions [10-11]. However, behind the user-friendly interfaces and engaging content lies a vast ecosystem of data collection processes that power these platforms. Figure: 2 depict the architectural model for sentiment analysis on women's abuse.

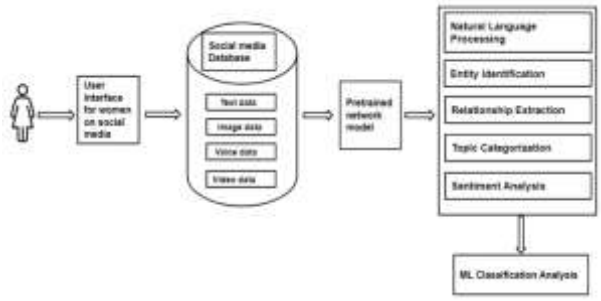

Figure: 2 Architectural Models for Sentiment Analysis on Women Abuses.

For this number of NLP libraries are available, including NLTK, spaCy, and TextBlob. To categorize the text data into distinct sentiment categories, such as positive, negative, or neutral, use a sentiment analysis algorithm [12]. It creates appropriate graphs or charts to represent the analysis of the sentiment data. This can aid in understanding the text data's general attitude towards women. Collecting and analyzing sentiment data related to women abuse is an essential step in understanding the public's perception and awareness of this critical issue [13]. Sentiment data can be gathered through surveys, social media monitoring, sentiment analysis of news articles and online discussions, and other data sources. Using modeling approaches, determine the many themes or subjects that can be found in the text data. This might assist to identify specific criticisms of women and the feelings that go along with them. Finally, the model analyses the observations and derives conclusions from them [14].

A. User Interface for women on Social Media Platforms: The discrimination towards women in social media has made it quite questionable for many people. Women are often objectified, sexualized, and commented in social media platforms in huge quantities. This can lead to unwanted attention, comments, and messages from people who view women as objective desire to fulfill their exhilaration. It’s important to recognize that women are in the problem, and are targeted with harmful behaviors on social media [15-16]. The role of women protection in social media platforms is crucial in ensuring a safe and inclusive online environment requires proper user interface to create safety for women users. Women often face various forms of harassment, abuse, and discrimination on social media, making it essential for these platforms to take proactive measures to protect their rights and well-being. Here are some key aspects of the role of women protection in social media to be protected [17-18].

B. Social Media Database: A social media database's scalability is defined as its capacity to manage higher data volumes, accommodate more concurrent users, and preserve optimal performance as the platform expands. Given the enormous and constantly expanding volume of data produced by social media platforms and the requirement to support an ever-increasing number of users, scalability is a crucial component of a social media database [19-20]. Social media platforms generate an enormous amount of data including text, images, videos, and audio. Social media platforms are a rich source of sentiment data where users often communicate their emotions and opinions in their posts and comments [21]. Social media databases are used to train learning models that can analyze the sentiment of a piece of text, which understanding both Named entity recognition and Language Modeling [22-23]. These tweets datasets are available on a large scale with diverse sources of language at various data repositories like Kaggle and Github etc. This dataset can be used to train machine learning or deep learning network model and improve the accuracy using NLP techniques. Social media databases can be used in a wide variety of ways in NLP research [24-25]. There are numerous datasets available for social media across various platforms. Few examples are described in the following Table: 1.

Table: 1 Description of Social Media Databases.

Social Media Platforms	Feature and Descriptions
Twitter API	Developers can access a multitude of data from Twitter, such as tweets, user profiles, and trends, through the public API that Twitter makes available.
Facebook Graph API	User access to Facebook and public data, which includes user profiles, pages, groups, events, and more, is made possible by the Facebook Graph API.
YouTube Data API	Developers can access video, channel, and playlist data from YouTube using the YouTube Data API.
Reddit API	Developers can get data on posts, comments, and subreddits by means of the Reddit API.
Instagram API	Instagram offers developers access to user profiles, postings, and comment data through an API.
LinkedIn API	User can access to information on user profiles, business profiles, job postings, and more is possible through the LinkedIn API.

Results and Discussion:

In order to implement and evaluate the NLP with machine learning methods successfully, it is essential to understand the mathematical and statistical methods for enhanced evaluations. Each NLP with machine learning technique has its own distinct behavior and can be applied to various types of research challenges for more accurate assessments. This research study is related to various research papers published on women's harassment from different countries. The detailed research review analysis is conducted on the related published work along with search results used in the relevant studies and comparisons. This model assessment is used to validate the datasets and supporting technology platforms with accuracy and acknowledged its quality results. The reviewed quality studies evaluated some synthesized findings on women's criticisms on various social media platforms. This model of study summarizing the key findings like Techniques-Wise study, Dataset-wise studies, and Platform-wise from relevant references papers are shown in detail below in Table: 2.

Table: 2 Represents the Related Work and Analysis Reports.

Sno	Reference Papers	Technical Algorithms	Tools	Datasets	Accuracy
1	[7] 2018	SVM	Python	Twitter	78.4%
2	[2] 2019	SVM	MIME	Meme	76%
3	[9] 2019	LSTM + Glove+ SVM	Python	Kaggle	98%
4	[3] 2020	SVM	Python	Twitter	89.9%
5	[4] 2020	M-BERT	BERT	Twitter	74%
6	[14] 2020	Naïve Bayes	Python	Kaggle	93%
7	[17] 2020	Naïve Bayes	Python	Twitter	80.62%
8	[18] 2020	RF	Python	Twitter	93%
9	[19] 2020	SVM	Python	Twitter	98.71%
10	[5] 2021	K-Mean	Python	Twitter	90%
11	[21] 2021	SVM	Python	Twitter-Arabic	86.89%
12	[1] 2022	SVM	Python	Synthesized Keywords.csv	98%
13	[6] 2022	BERT	Python	Github	90.43%
14	[25] 2022	Naïve Bayes	Python	Weka	85.42%

Most of the research works outperformed using NLP with machine learning techniques with different accurate results. The Analysis report of the research paper results is shown in Figure: 3.

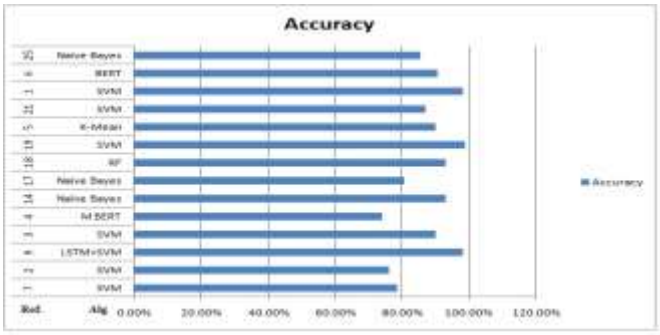

Figure: 3 Analysis Report of the Research Paper.

Conclusion: Most of the women's criticism and harassment are identified through negative speech which is analyzed, and taken into consideration such as negative slang, negative emotion, negative negation expression, and emotional dictionaries which are carrying negative remarks on women in social media platforms. Based on past surveys women's criticism papers related to NLP with ML algorithms are supported for better understanding to enhance present work. Unwanted text can easily be identified using some supervised and unsupervised techniques based on text classification. In the given reference paper [19] 2020 the SVM-supervised technique practical on the Twitter dataset has achieved the best accurate result of 98.71% in past research-related works. Next reference paper [14] 2020 researcher implemented Naïve Bayes supervised techniques applied on the Kaggle dataset has achieved the best accurate result of 93% achieved effectively.

Future Scope: Compared to the current system, the outcome will be more efficient. The next extended work of the new algorithm for classifying text is based on the negative emotion opinion and emoji expressed by the users in social

media platforms are taken into the consideration. By considering the present research survey SVM techniques and the BERT model are suggested for future research work and development. This research can help women who are facing various kinds of troubles on social media platforms by protecting their self-respect and politeness.

References:

[1] https://www.theatlantic.com/technology/archive/2014/10/the-unsafety-net-how-social-media-turned-against-women/381261/
[2] Alkomah, F., & Ma, X. (2022). A literature review of textual hate speech detection methods and datasets. Information, 13(6), 273.
[3] Jahan, M. S., & Oussalah, M. (2023). A systematic review of Hate Speech automatic detection using Natural Language Processing. Neurocomputing, 126232.
[4] Masruroh, S. U., Utami, D. Z. A., Khairani, D., Azhari, M., Helmi, M. I., & Putri, R. A. (2022, September). Sentiment Analysis on Twitter towards the Ratification of a Bill on the Elimination of Sexual Violence in Indonesia using Machine Learning. In 2022 10th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-5). IEEE.
[5] Fersini, E., Gasparini, F., & Corchs, S. (2019, September). Detecting sexist meme on the web: a study on textual and visual cues. In 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW) (pp. 226-231). IEEE.
[6] Madhubala, D., Rajendiran, M., & Elangovan, D. (2020, November). A Study on Effective analysis of Machine Learning algorithm towards the Women's safety in Social Media. In 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1151-1156). IEEE.
[7] Rodríguez-Sánchez, F., Carrillo-de-Albornoz, J., & Plaza, L. (2020). Automatic classification of sexism in social networks: An empirical study on twitter data. IEEE Access, 8, 219563-219576.
[8] González, G. A. R., & Cantu-Ortiz, F. J. (2021, March). A Sentiment Analysis and Unsupervised Learning Approach to Digital Violence Against Women: Monterrey Case. In 2021 4th International Conference on Information and Computer Technologies (ICICT) (pp. 18-26). IEEE.
[9] Soldevilla, I., Flores, N., & Tuesta, S. (2022, May). Macht: An Application Based on a Sentiment Analysis Model to the Identification of Messages in Spanish with Gender Violence Content on Twitter. In 2022 5th International Conference on Artificial Intelligence and Big Data (ICAIBD) (pp. 396-401). IEEE.
[10] Watanabe, H., Bouazizi, M., & Ohtsuki, T. (2018). Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection. IEEE access, 6, 13825-13835.
[11] Nagamanjula, R., & Pethalakshmi, A. (2018, June). A Machine Learning Based Sentiment Analysis by Selecting Features for Predicting Customer Reviews. In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1837-1843). IEEE.
[12] Sajjad, M., Zulifqar, F., Khan, M. U. G., & Azeem, M. (2019, August). Hate speech detection using fusion approach. In 2019 International Conference on Applied and Engineering Mathematics (ICAEM) (pp. 251-255). IEEE.
[13] Singh, P., Pisipati, N., Krishna, P. R., & Prasad, M. V. (2019, February). Social signal processing for evaluating conversations using emotion analysis and sentiment detection. In 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP) (pp. 1-5). IEEE.
[14] Noor, A., & Islam, M. (2019, July). Sentiment Analysis for Women's E-commerce Reviews using Machine Learning Algorithms. In 2019 10th International conference on computing, communication and networking technologies (ICCCNT) (pp. 1-6). IEEE.
[15] Ansari, R., Jaffar, B. A., Riaz, S., Kaur, M. J., & Mushtaq, A. (2019, February). Datamining to alert the formation of women objectification stereotypes in video games. In 2019 Amity International Conference on Artificial Intelligence (AICAI) (pp. 521-526). IEEE.
[16] Tan, F., Liu, W., & Zhao, X. (2020, November). Effect of Sexism and Self-Esteem Toward Self-Objection. In 2020 International Conference on Public Health and Data Science (ICPHDS) (pp. 64-67). IEEE.

[17] Ullah, M. A., Hasnayeen, M. A., Shan-A-Alahi, A., Rahman, F., & Akhter, S. (2020, December). A Search for Optimal Feature in Political Sentiment Analysis. In 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE) (pp. 340-343). IEEE.
[18] Yu, J., Aduragba, O. T., Sun, Z., Black, S., Stewart, C., Shi, L., & Cristea, A. (2020, August). Temporal Sentiment Analysis of Learners: Public versus Private Social Media Communication Channels in a Women-in-Tech Conversion Course. In 2020 15th International Conference on Computer Science & Education (ICCSE) (pp. 182-187). IEEE.
[19] Li, G., Zheng, Q., Zhang, L., Guo, S., & Niu, L. (2020, November). Sentiment infomation based model for Chinese text sentiment analysis. In 2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE) (pp. 366-371). IEEE.
[20] Kaur, C., & Sharma, A. (2020, October). Social issues sentiment analysis using python. In 2020 5th International Conference on Computing, Communication and Security (ICCCS) (pp. 1-6). IEEE.
[21] Herodotou, H., Chatzakou, D., & Kourtellis, N. (2020, December). A streaming machine learning framework for online aggression detection on twitter. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 5056-5067). IEEE.
[22] Rini, R., Utami, E., & Hartanto, A. D. (2020, October). Systematic Literature Review Of Hate Speech Detection With Text Mining. In 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS) (pp. 1-6). IEEE.
[23] Fareed, S. B. S. (2021, March). API Pipeline for Visualising Text Analytics Features of Twitter Texts. In 2021 International Conference of Women in Data Science at Taif University (WiDSTaif) (pp. 1-6). IEEE.
[24] Alzyout, M., Bashabsheh, E. A., Najadat, H., & Alaiad, A. (2021, May). Sentiment Analysis of Arabic Tweets about Violence Against Women using Machine Learning. In 2021 12th International Conference on Information and Communication Systems (ICICS) (pp. 171-176). IEEE.
[25] Hakami, S. A. A., Hendley, R., & Smith, P. (2022, December). Gender Impact on Emoji Sentiment Analysis among Arabic Users in Digital Networks. In 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 199-203). IEEE.

Bibliography:
1. Henry, N., & Powell, A. (2018). "Embodied harms: gender, shame, and technology-facilitated sexual violence." Feminist Media Studies, 18(4), 543-558.
2. Amnesty International. (2018). Report on "Toxic Twitter: A Toxic Place for Women."
3. Citron, Danielle Keats, "Hate Crimes in Cyberspace" (2014). Book Gallery. 91. https://digitalcommons.law.umaryland.edu/books/91
4. Pew Research Center. (2017). "Online Harassment 2017."
5. United Nations Broadband Commission for Digital Development. (2015). "Cyber Violence Against Women and Girls: A Worldwide Wake-Up Call."