Data Analytics on Twitter for Evaluating Women Inclusion and Safety in Modern Society

Loredana Caruccio, Stefano Cirillo, Vincenzo Deufemia, Giuseppe Polese and Roberto Stanzione

University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, Salerno, Italy

Abstract

The inclusion and safety of women in modern society is a problem that over the years has become increasingly important in all countries, leading to a large number of awareness campaigns and social movements. Thus, several research communities have investigated various aspects of this problem, mainly focusing on interpreting the behaviors of people, also according to their geopolitical situations, for preventing and reducing discrimination situations against women. In this paper, we present a study on the resonance of the gender equality issues on the Twitter social network. In particular, we collected 4.4 million tweets shared by over 1 million users and applied data analytics techniques to evaluate statistics, correlations between the most common hashtags, and location-based sentiment analysis.

Keywords

Big Data Analytics, Gender Equality, Woman Inclusion, Twitter, Sentiment Analysis

> "*I raise up my voice – not so that I can shout, but so that those without a voice can be heard*".
>
> Malala Yousafzai

1. Introduction

The inclusion of women in society represented one of the main challenge that accompanied the history in the last centuries. In fact, although women always contributed to the cultural, scientific, and economic growth of the society, they had to fight for obtaining own rights and for starting in being considered equal to the men. The major examples of women's battles are related to suffrage campaigns, which gave women the right of voting only in the last century (in most cases). Nevertheless, although from a legal point of view all women have the right to vote, in some regions appear quite difficult for them to apply their right due to societal norms, harassment, and violence at the polls, or pressure from their husbands[1].

The gender equality still represents an important achievement to be addressed in the modern society. In fact, it has been included into the 17 Sustainable Development Goals defined by the United Nations (UN) in 2015[2]. Within it, the women's empowerment and gender equality have been structured as 9 specific worldwide targets that must be reached until 2030, which range

[1]https://worldpopulationreview.com/country-rankings/countries-where-women-cant-vote
[2]https://www.globalgoals.org/goals/5-gender-equality/

from the ending of discrimination and violence against women to the inclusion of women in political, economical, and decision-making processes (by also promoting it through technology). In fact, different studies highlighted that the gender equality not only represents a human right issue, but also a precondition for the sustainable development [1], even though gender inequalities across economic, social and environmental dimensions remain widespread and persistent, especially in some areas of the world [2]. Consequently, many social movements and organizations were born with the mission of promoting the gender equality and highlighting any discrimination towards the woman empowering in the society. They typically feed the worldwide resonance of the problem, enabling people to be informed, think, and act against it.

In this paper, we present the analysis that we performed on the resonance of the gender equality issues over the Twitter social network. In particular, we collected 4,4 million tweets shared by over 1 million users across the world, in a topic-specific dataset, named WISHES-Women Inclusion and Safety for Holding Equality in Society. Furthermore, after an hydrating process, the dataset has been first used to perform a quantitative analysis highlighting major statistics of collected tweets, which have been successively used to perform a sentiment analysis by also considering the geopolitical scenario of different countries.

The remainder of the paper is organized as follows. In Section 2 we describe relevant works concerning previous analysis over the gender equality issues. Then, Section 3 introduces the WISHES dataset, by highlighting the building process we considered, and the main statistics concerning collected tweets. Instead, the evaluation of the people perception, performed through a sentiment analysis, is presented in Section 4. Then, final considerations and future directions conclude the paper.

2. Related Work

Over the past decade, social media platforms have been an important means of communication for people, and their impact on daily lives is increasing every day. Thanks to their accessibility, social networks represent a big source of information that can be used to extract insights for analyzing collective thinking about specific issues, such as the role of women in the modern society. In this section, we provide an overview of some related works dealing with the issues of violence against women and the digital gender divide, which represent the main topics we considered throughout the proposed analysis.

In [3] the authors analyzed 2,5 million tweets and then compared tweets about VAW (Violence Against Women) with tweets on other topics like politics or sports. In particular, they discovered that for the VAW topic there is a higher average of users and tweets per thread and also a higher average of thread depth. These tweets have been further analyzed in [4], in which authors focused their study on tweets in Arabic, by performing a sentiment analysis to extract insights on how the issue of VAW is treated, and they noticed that this topic is gaining attention in the Arabic world. Similar analysis also focused to South African [5] and Indonesian tweets [6].

In other recent studies, authors have extracted and analyzed the most used terms and the most discussed topics related to domestic violence in the shared tweets [7] and in news articles [8], by mainly focusing the analysis on articles related to gender-based violence from newspapers in India, Pakistan, and the United Kingdom. In [9], the authors conducted a similar analysis in Australia, considering live chats on Twitter related to a special *Q&A* episode dedicated to

the topic of domestic violence. The authors collected tweets written from the beginning of the episode until the next day, and the results of their study showed that there were strongly conflicting opinions and no agreement on condemning domestic violence. Other Recent articles, such as [10] and [11], investigated the possible correlation between misogynistic tweets and episodes of violence in the United States. In particular, the first study reported a correlation with the number of rapes per capita, while the second one highlighted a correlation between tweets and episodes of domestic violence.

Another topic on which the research has focused is the digital gender divide (DGD), which indicates imbalances in access to Information and Communication Technologies (ICTs). Several studies focus on developing countries [12, 13, 14], highlighting the obstacles faced by women in the use of technology. In these studies, authors mention lack of education, unfavorable employment, and income conditions as main causes of DGD.

As widely discussed in [15] and [16], the DGD problem does not exist only in developing countries. In fact, these studies have highlighted how this disparity is also present in Europe. Among the mentioned factors, there are stereotypes, prejudice, male hostility, and other women-related problems like pregnancy and maternity. Quantifying DGD is not always easy because of the lack of data, especially in low-income countries. To address this problem, in [17] the authors proposed a method to track the gender digital gap using Facebook's advertisement audience estimates. They showed that with the proposed method it is possible to extend the calculation of the gender gap indicators to more countries, especially low and middle-income ones.

Unlike the papers introduced above, the proposed study introduces a new up-to-date dataset of tweets related to the role of women in different geopolitical areas, and provides a detailed multilingual sentiment analysis to evaluate the inclusion and safety of women in modern society.

3. The wishes dataset

In this section, we provide an overview of the wishes dataset and analyze the different types of tweets collected from Twitter. Moreover, we discuss the strategy for retrieving the content of tweets with the aim to extract statistics from them and analyze the correlation between the most frequent hashtags. Finally, we provide a preliminary analysis of people's collective thoughts by analyzing the content of the tweets. wishes dataset is publicly accessible on GitHub[3].

3.1. Overview of the wishes Dataset

The creation of the dataset has required the definition of an extraction module capable of continuous monitoring tweets shared by users and collecting the most pertinent ones. This module has been configured for interacting with Twitter streaming API and reading the tweets related to a set of keywords concerning the role of women in the society and domestic violence.

To define the set of keywords involved in our monitoring, we started from a small set of popular keywords manually extracted from the most popular hashtags used on Twitter that have been provided by online services[4], and collected tweets for one day aiming to extract keywords that co-occur with the initial ones. We have considered a set over 30 keywords representing

---

[3]https://github.com/Roby46/Wishes_tweets_dataset
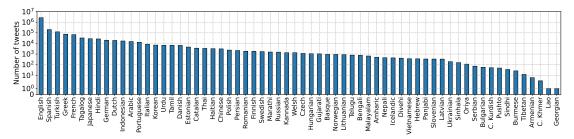[4]Tweeplers

Figure 2: Language distribution of tweets in WISHES dataset.

the most used in tweets from around the world, such as "*#abuseisnotlove*", "*#youarenotalone*", "*#domesticabuse*", or "*#domesticviolence*".

The collection step has been continuously performed since November 10th, 2021, and by February 25th, 2022, achieving about 4.4 million tweets shared by over 1 million users. According to Twitter policies, during the monitoring steps, we stored only the IDs of the tweets, requiring the consequent hydration of them for analyzing their contents. In particular, the hydration process of tweets has requested to connect with the Twitter API for accessing to the content of the tweets. Table 1 shows statistics on tweets extracted before and after the hydration process. As we can see, the number of tweets has been reduced to 3,760,574, from which 17,537 are written by verified users, whereas 1,170,868 represent unverified tweets. This could be due to the fact that either the tweets were written by users who have been banned on Twitter, or they represent tweets that have been removed from the platforms.

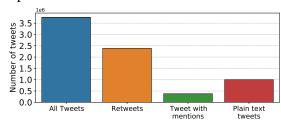|	Before Hydrating	After Hydrating
Number of tweets	4,489,070	3,760,574
Number of accounts	-	1,170,868
Verified accounts	-	17,537
Accounts with location	-	2,507
Average tweets per account	-	3.21
Most recent tweet	2022-02-25	

Table 1: Details of WISHES dataset.

Figure 1: Types of tweet in WISHES dataset.

After hydrating the content of the tweets, we performed a cleaning process by removing all special characters and/or emojis, yielding the standardization of the tweet syntax. This operation allowed us to identify the most common hashtags used in tweets and analyze their frequency in the dataset.

Figure 1 shows the different types of tweets shared by both verified and non-verified accounts [18]. As we can see, among the different types of tweets, there are about 2.5 millions of retweets (66% of all tweets), about 500,000 of tweets contain at least one mention towards other users (13% of all tweets), and about 1 million of tweets are *plain text tweet* (26% of all tweets), i.e., tweets written in a standard text-based format that are not retweets and do not contain mentions. Figure 2 shows the number of tweets for each language collected in WISHES. As we can see, WISHES contains tweets written in 64 different languages, most of which are written in English, Spanish, and Turkish.

In order to perform a more detailed analysis of the collective thinking on the role of women in society and on the problem of domestic violence, we provide a preliminary analysis of the

most frequent hashtags used in the tweets as shown in Figure 3. As we can see, among the most frequent hashtags (i.e., the hashtags with the largest font sizes), there are hashtags that seem to not be related to the purpose of the proposed study, such as "*#quoteofthedays*" and "*#endsars*". In fact, although the former is a generic hashtag used by Twitter users when writing a sentence for the day, the latter refers to a decentralized social movement and series of mass protests against police brutality in Nigeria. Nevertheless, in both cases, the tweets including these hashtags also contain references to the woman and their roles in society. Furthermore, it is possible to notice that most hashtags have targeted references closer to the scope of WISHES, such as "*#metoo*", "*#domesticviolence*", "*#selflove*", "*#heforshe*", and "*#neverforget*".

Table 2 shows the occurrences and the frequencies of the top 30 hashtags used in the collection of tweets. The frequency is the number of occurrences of each hashtag with respect to the tweets shared from both verified and non-verified accounts. It is important to notice that only 2,741,802 of 3,760,574 tweets contain at least one hashtag, which corresponds to 73% of all the tweets. As shown in Table 2, we provide three different values representing the frequency of each hashtag with respect to *i*) the number of all the hashtags used in the tweets (i.e., N_X); *ii*) the number of tweets that contain at least one hashtag (i.e., N_Y), and *iii*) the number of all tweets in the data collection (i.e., N_Z). In particular, let N be the number of occurrences of each hashtag in the data collection, the frequencies are defined as follows:

$$F_1 = \frac{N \cdot 100}{N_X} \qquad F_2 = \frac{N \cdot 100}{N_Y} \qquad F_3 = \frac{N \cdot 100}{N_Z} \qquad (1)$$

From Table 2 we can notice that, there is a large number of tweets containing hashtags that refer to social movements against abuse and solidarity movements for the advancement of gender equality, such as "*#metoo*", "*#niunamenos*", "*#orangetheworld*", and "*#heforshe*". Among them, one of the most frequent hashtags is "*#endsars*", which, as said above, is used for referring to a series of demonstrations against the police that is accused of torture and other crimes in Nigeria. However, the use of these hashtags in most tweets shows that these tortures and crimes also affect Nigerian women. Moreover, we can notice that these hashtags are contained in approximately 30.82% of the tweets of users who have used at least one hashtag in their posts, and 22.47% of the entire collection of tweets.

Many of the hashtags used refer to domestic life scenarios, which still today represent an important social scourge, such as "*#relationships*", "*#divorce*", "*#domesticviolence*", and "*#never-*

Figure 3: Hashtags used in WISHES dataset.

Hashtag	Occurrences	F_1(%)	F_2(%)	F_3(%)	Hashtag	Occurrences	F_1(%)	F_2(%)	F_3(%)
#endsars	375,592	3.56	13.70	9.99	#women	63,063	0.60	2.30	1.68
#metoo	216,588	2.05	7.90	5.76	#niunamenos	59,542	0.56	2.17	1.58
#survivor	194,176	1.84	7.08	5.16	#heforshe	57,912	0.55	2.11	1.54
#neverforget	186,327	1.77	6.80	4.95	#inspiration	51,268	0.49	1.87	1.36
#quoteoftheday	186,220	1.77	6.79	4.95	#mentalhealth	44,849	0.43	1.64	1.19
#children	152,653	1.45	5.57	4.06	#divorce	40,350	0.38	1.47	1.07
#selflove	103,229	0.98	3.77	2.75	#education	37,427	0.35	1.37	1.00
#positivevibes	99,874	0.95	3.64	2.66	#therapy	37,048	0.35	1.35	0.99
#healing	92,818	0.88	3.39	2.47	#domesticviolence	36,875	0.35	1.34	0.98
#love	78,735	0.75	2.87	2.09	#selfcare	36,174	0.34	1.32	0.96
#orangetheworld	75,034	0.71	2.74	2.00	#life	35,900	0.34	1.31	0.95
#quote	71,699	0.68	2.62	1.91	#survivor2022allstar	35,773	0.34	1.30	0.95
#motivation	65,138	0.62	2.38	1.73	#business	35,238	0.33	1.29	0.94
#relationships	64,552	0.61	2.35	1.72	#survivorallstar2022	34,151	0.32	1.25	0.91
#quotes	64,461	0.61	2.35	1.71	#video	34,050	0.32	1.24	0.91

Table 2
Top 30 hashtags in WISHES dataset.

forget". These hashtags are contained in approximately 4.42% of the tweets of users who have used at least one hashtag in their posts, and 3.22% of the entire collection of tweets. These preliminary statistics allow us to estimate how much domestic violence and gender inequality have sparked strong thoughts in people around the world.

3.2. Correlation Analysis

Among the different kinds of correlation that can be analyzed [19], we used Spearman's Rank-Order Correlation [20] to evaluate the correlation between hashtags, by applying it to the top-50s hashtags appearing in at least 30,000 different tweets. The values obtained from this analysis are comprised in the range $[-1, 1]$. In particular, given two elements x and y, a zero correlation value implies no relationship between them, while values -1 and 1 indicate monotonic positive and negative relationships, respectively. More specifically, a positive correlation means that the increase of x leads to the consequent increase of y, whereas a negative correlation means that when x increases, y decreases. Two elements have a high correlation value if they have a similar rank, whereas if their rank is significantly different they have a low correlation value.

Figure 4 shows the correlation matrix of the most frequent hashtags used in the tweets of WISHES. As we can see, there is no evidence of strong negative correlations between hashtags. However, there are some weak correlations between "*#children*" and the hashtags "*#quoteofthe-day*" and "*#metoo*". In the first case, the negative correlation is probably due to the fact that the tweets containing quotes of the day often do not contain any reference to children, but they are reflexive aphorisms about woman and their role in society. In the second case, the negative correlation between "*#children*" and "*#metoo*" is due to the fact that the *metoo* movement is mainly about harassment in the workplace, leading it to be not significantly related to children.

Other significant correlations can be found between "*#motivation*", "*#mindset*", with "*#motivationalquotes*", and between "*#inspiration*" with "*#inspirationalquotes*". These positive correlations indicate that many tweets and users provide motivational messages for women living in distressed family and work situations.

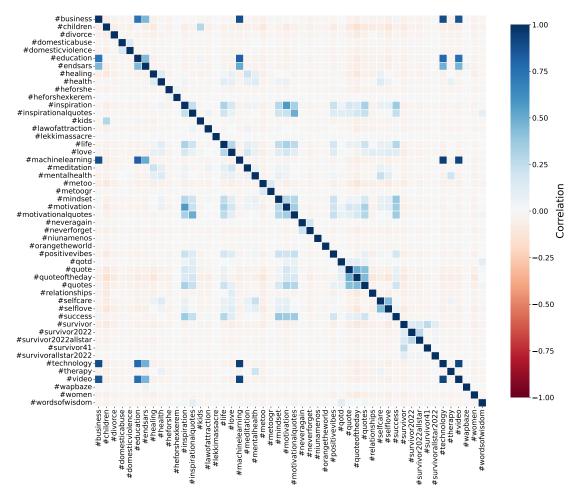Another topic covered by tweets in WISHES dataset regards the relationship between the

Figure 4: Correlation matrix of the top-50 hashtags that appear in at least 30.000 different tweets.

women and their children. In particular, we can notice that there exist moderate correlations between the hashtags "*#children*" with "*#kids*" and "*#education*" meaning that when tweets discuss children, it is often women who express thoughts about the relationship of them with parents and with the school.

Finally, other interesting correlations can be found between the hashtags "*#healing*" and "*#mentalhealth*", and between "*#selfcare*" and "*#selflove*". These hashtags are contained in tweets in which many people ask for support and positive opinions for healing from situations of violence that have indissolubly marked their lives. Moreover, they incite self and free love without borders, because, in modern society, situations of discrimination and violence capable of limiting people's freedom are no longer acceptable.

4. Sentiment Analysis

Sentiment analysis is a field of natural language processing (NLP) that allows to automatically evaluate user comments in order to determine whether a text expresses a positive or negative opinion. It is based on the main methods of computational linguistics and textual analysis,
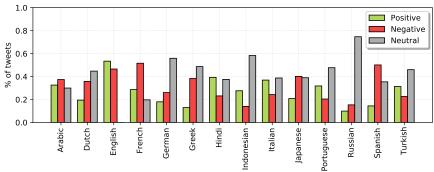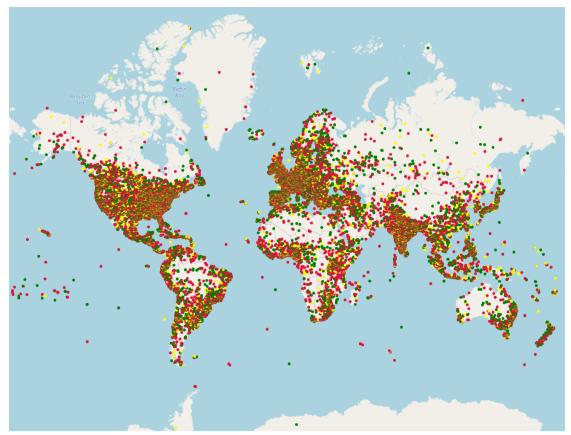
Figure 5: Sentiments detected for the analyzed languages.

and it is adopted in several sectors, such as politics [21], stock markets [22], marketing [23] and sports [24]. In our study, the sentiment analysis aims at evaluating the feelings expressed by users about gender inclusion and the role of women in different countries. In particular, we performed a multi-language sentiment analysis by considering 3,275,529 tweets written in Arabic, Dutch, English, French, German, Greek, Hindi, Indonesian, Italian, Japanese, Portuguese, Russian, Spanish, and Turkish languages. The sentiment analysis on all these languages has been performed using *XLM-Roberta* [25]. This model is based on Google BERT (Bidirectional Encoder Representations from Transformers) [26], which is a bidirectional model capable of identifying the meaning of words by considering only the other words contained in a sentence. BERT has been designed to be easily fine-tuned for multiple tasks, like language inference and it has been already trained on a huge number of text in different languages.

Figure 5 shows the sentiment percentages detected for each languages involved in our evaluation. As we can see, tweets written in Hindi and English represent the only ones where the percentage of positive sentiment exceeds both neutral and negative sentiments, i.e., 39.32% and 53.42%, respectively. Moreover, we can notice that for Russian, German, Dutch, and Greek languages there is a predominance of neutral sentiment and low percentages of positive tweets, i.e., 9.9%, 18%, 19%, and 12.9%, respectively. Concerning Italian, Portuguese, Indonesian, and Turkish tweets they present a predominantly neutral sentiment, but the percentage of positive tweets exceeds the negative ones. Furthermore, we can notice that for Spanish, Japanese, French, and Arabic tweets the main sentiments is negative. In fact, French tweets have the highest percentage of negative sentiment, i.e., 51.62%, whereas for Arabic tweets the percentages of tweets is similar for all different types of sentiments.

Although the previous analysis provides an overview of the sentiments of the tweets for each language, to analyze the impact of the sentiments in relation to the geopolitical situation of the various countries, we performed a more detailed analysis of the tweets by taking into account the geographical origin of the tweets (Figure 6). Positive tweets are shown in green, negative tweets in red, and in yellow the neutral ones.

As we can see, there is a high concentration of tweets from the American continent, Indonesia, Central Europe, and India. There is also a relevant number of tweets from China, Japan, and some African countries, such as Nigeria, South Africa, Kenya, and Somalia. In particular, the highest concentration of negative tweets is from India, Africa, and some areas of South America. A possible explanation is that in these countries there is an important part of the

Figure 6: Analysis of the geographical origin of tweets in WISHES dataset.

population living in difficult conditions and there still exists a strong disparity between women and men, compared to more developed countries. In fact, Jewkes highlighted how poverty and the associated stress are key factors that contribute to violence and that affect women's empowerment [27]. Considering the countries with many negative tweets, we can indeed notice that often the living conditions are difficult. The economic and political situation of African countries has led this continent to be classified as the last in terms of nominal gross domestic product per capita, according to the International Monetary Fund[5]. Similarly, also South America suffers from the problem of poverty and it is the second to last continent in this ranking. Concerning India, overpopulation can be another negative factor, because it leads to high levels of unemployment and unfair income distributions. This generates stress and discontent, which can lead to violent behaviors. In fact, as reported from the *National Family Health Survey-5*[6], nearly 30% of women in India have experienced sexual or physical violence.

Moreover, the *Women Peace & Security Index*[7] have ranked 170 countries based on women's inclusion and safety, by taking into account several factors, such as social norms, education levels, and economic situations. Among all these countries, it is possible to notice that countries

---

[5]https://statisticstimes.com/economy/continents-by-gdp-per-capita.php
[6]http://rchiips.org/nfhs/NFHS-5_FCTS/India.pdf
[7]https://giwps.georgetown.edu/wp-content/uploads/2021/11/WPS-Index-2021.pdf

like India, Nigeria, Somalia, and Congo are in rather backward positions, i.e., 148th, 130th, 159th, and 155th position, respectively. Furthermore, it is important to notice that the presence of the *endsars* movement in Nigeria could contribute to more negative tweets, since they highlight the violence of the police against all the Nigerian people.

Concerning the Central Europe, we can notice that there is a greater balance between positive and negative tweets. The latter are probably due to the episodes of violence that probably raised reactions by several negative discussions on Twitter. Instead, the high number of positive tweets is probably due to the fact that the awareness campaigns and movements, such as *MeToo* and *HeForShe*, are positively impacting the European population. Similarly, for the population of Latin America, the spread of the awareness campaign of the *Niunamenos* movement has probably led to a greater number of positive tweets.

The analysis discussed in this section revealed interesting differences in the sentiments expressed in tweets with respect to the language and geographic origin of tweets. Unfortunately, there are still many situations and countries in the world that require more intensive awareness campaigns on violence against women and in favor of gender equality in order to sensitize different cultures to the criteria of non-violence and respect for other people.

5. Conclusion and Future Directions

Aiming at analyzing the resonance of topics related to the inclusion and safety of women in modern society, with this study we collected a huge quantity of tweets concerning gender equality and domestic violence in the wishes dataset (also shared in a public repository). Then, we analyzed the collected tweets by highlighting the most used hashtags and the correlations between them. Finally, we performed a sentiment analysis process in order to emphasize the positive/negative inflection of tweets with respect to the tweet's language and/or geopolitical areas from which they have originated. The performed analysis allowed us to notice that many of the collected tweets are related to no-profit movements and/or awareness campaigns, which promote woman empowerment and promptly highlight violent episodes according to events occurring throughout the world. An interesting witness of the proposed analysis is related to the tweet concentration (also according to sentiments they include), which revealed a different perception with respect to the social and political scenario in countries.

Starting from this preliminary analysis, many other deep analysis could be performed. For instance, it is possible to analyze contents and sentiments of tweets according to features more related to the user profiles and their influence over social networks. These analyses could be performed also exploiting topic modeling techniques to more accurately identify the content of tweets with the aim to obtain a more detailed analysis of them. Moreover, studies could be conducted by specifically considering events and/or awareness campaigns. Finally, a comparative analysis could also be accomplished by collecting and considering posts from other social networks, blogs, and websites [28].

References

[1] M. L. Alvarez, From unheard screams to powerful voices: A case study of women's political empowerment in the philippines, in: 12th National Convention on Statistics, 2013, pp. 1–73.

[2] E. Bayeh, The role of empowering women and achieving gender equality to the sustainable development of ethiopia, Pacific Science Review B: Humanities and Social Sciences 2 (2016) 37–42.

[3] J. Xue, K. Macropol, Y. Jia, T. Zhu, R. J. Gelles, Harnessing big data for social justice: An exploration of violence against women-related conversations on twitter, Human Behavior and Emerging Technologies 1 (2019) 269–279.

[4] M. Alzyout, E. A. Bashabsheh, H. Najadat, A. Alaiad, Sentiment analysis of arabic tweets about violence against women using machine learning, in: 2021 12th International Conference on Information and Communication Systems (ICICS), IEEE, 2021, pp. 171–176.

[5] J. Oyasor, M. Raborife, P. Ranchod, Sentiment analysis as an indicator to evaluate gender disparity on sexual violence tweets in south africa, in: 2020 International SAUPEC/Rob-Mech/PRASA Conference, IEEE, 2020, pp. 1–6.

[6] K. Budiman, N. Zaatsiyah, U. Niswah, F. M. N. Faizi, Analysis of sexual harassment tweet sentiment on twitter in indonesia using naïve bayes method through national institute of standard and technology digital forensic acquisition approach, Journal of Advances in Information Systems and Technology 2 (2020) 21–30.

[7] J. Xue, J. Chen, R. Gelles, Using data mining techniques to examine domestic violence topics on twitter, Violence and gender 6 (2019) 105–114.

[8] M. A. Manzoor, S.-U. Hassan, A. Muazzam, S. Tuarob, R. Nawaz, Social mining for sustainable cities: thematic study of gender-based violence coverage in news articles and domestic violence in relation to covid-19, Journal of ambient intelligence and humanized computing (2022) 1–12.

[9] M. Dragiewicz, J. Burgess, Domestic violence on# qanda: The "man" question in live twitter discussion on the australian broadcasting corporation's q&a, Canadian journal of women and the law 28 (2016) 211–229.

[10] R. Fulper, G. L. Ciampaglia, E. Ferrara, Y. Ahn, A. Flammini, F. Menczer, B. Lewis, K. Rowe, Misogynistic language on twitter and sexual violence, in: Proceedings of the ACM Web Science Workshop on Computational Approaches to Social Modeling, 2014, pp. 57–64.

[11] K. R. Blake, S. M. O'Dean, J. Lian, T. F. Denson, Misogynistic tweets correlate with violence against women, Psychological science 32 (2021) 315–325.

[12] M. Hilbert, Digital gender divide or technologically empowered women in developing countries? a typical case of lies, damned lies, and statistics, Women's Studies International Forum 34 (2011) 479–489.

[13] N. O. Alozie, P. Akpan-Obong, The digital gender divide: Confronting obstacles to women's development in africa, Development Policy Review 35 (2017) 137–160.

[14] A. Antonio, D. Tuffley, The gender digital divide in developing countries, Future Internet 6 (2014) 673–687.

[15] M. Perifanou, A. Economides, Gender digital divide in europe, International Journal of Business, Humanities and Technology 10 (2020) 7–14.

[16] M. López-Martínez, O. García-Luque, M. Rodríguez-Pasquín, Digital gender divide and convergence in the european union countries, Economics 15 (2021) 115–128.

[17] M. Fatehkia, R. Kashyap, I. Weber, Using facebook ad data to track the global digital gender gap, World Development 107 (2018) 189–209.

[18] L. Caruccio, D. Desiato, G. Polese, Fake account identification in social networks, in: 2018

IEEE international conference on big data (big data), IEEE, 2018, pp. 5078–5085.

[19] L. Caruccio, V. Deufemia, F. Naumann, G. Polese, Discovering relaxed functional dependencies based on multi-attribute dominance, IEEE Transactions on Knowledge and Data Engineering 33 (2020) 3212–3228.

[20] C. Spearman, The proof and measurement of association between two things, The American Journal of Psychology 15 (1904) 72–101.

[21] J. Ramteke, S. Shah, D. Godhia, A. Shaikh, Election result prediction using twitter sentiment analysis, in: 2016 international conference on inventive computation technologies (ICICT), volume 1, IEEE, 2016, pp. 1–5.

[22] V. S. Pagolu, K. N. Reddy, G. Panda, B. Majhi, Sentiment analysis of twitter data for predicting stock market movements, in: 2016 international conference on signal processing, communication, power and embedded system (SCOPES), IEEE, 2016, pp. 1345–1350.

[23] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, H. Mora, Managing marketing decision-making with sentiment analysis: An evaluation of the main product features using text data mining, Sustainability 11 (2019) 4235.

[24] S. Aloufi, A. El Saddik, Sentiment identification in football-specific tweets, IEEE Access 6 (2018) 78609–78621.

[25] F. Barbieri, L. E. Anke, J. Camacho-Collados, Xlm-t: A multilingual language model toolkit for twitter, arXiv preprint arXiv:2104.12250 (2021).

[26] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[27] R. Jewkes, Intimate partner violence: causes and prevention, The lancet 359 (2002) 1423–1429.

[28] F. Cerruto, S. Cirillo, D. Desiato, S. M. Gambardella, G. Polese, Social network data analysis to highlight privacy threats in sharing data, Journal of Big Data 9 (2022) 1–26.