Sentiment Analysis System for Mapping Hate Speech against Women in Social Media using GIS System

Mohammad N. Shatnawi¹ and Nawras Shatnawi²

¹ Akdeniz University

Abstract
This study aims to map hate speech against women in the Middle East using a Geographic Information System (GIS) and sentiment analysis, with the goal of identifying patterns. The hate speech terms that were utilized in the research were gathered from more than 3600 women in the study region, according to the data.

Index terms—python, GIS, sentiment analysis, hate speech, social media, tweeter, women.

1 Introduction
Social media is transforming the face of culture and communication in the world. Social media in the Middle East has dramatically altered the way we speak to our friends, and live our lives in general. The decentralized nature makes it a perfect place to create and exchange ideas, data, images, videos, art, music, and more for amateurs and professionals alike. In 2020, around 3.8 million people using social media platforms, which is over half the population of the planet. People spent 144 minutes per day on social media, which mean the average person will spend more than six years of their lives on social media. (Kanyi, 2020). There are several advantages to social media, such as individual communication, collaboration, promoting access to information through open sources, offering a democratic forum for alternative viewpoints that cannot be heard in the mass media, creating autonomous self-governing institutions, encouraging citizen journalism, and collective knowledge. (Fuchs, 2006). At the same time, there are several consequences such as online harassment, trolling, cyber-bullying, and hate speech which has been investigated in the current study. Although the term “hate speech” is widely used, there is no universally accepted definition for it. Most countries have adopted legislation prohibiting expressions according to the definitions of “hate speech” that differ slightly when defining what is prohibited (Synodinou et al, 2019). The term hate speech is understood as any kind of communication in speech, writing or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are (UN, 2019). Twitter defines hate speech as any tweet that promotes violence against other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease (Twitter, 2020) Although the terms of services of these platform typically forbid hateful and harassing speech, applying these rules still challenging, as automatically identifying hate speech is not enough measured (Wassem & Hovy, 2016). As we move into a moment of heightened scrutiny of social media platforms, their scope, and use, as well as the new technologies on which they are built, continue to expand. And it is ingrained in the daily lives of a large number of individuals. Despite the fact that social networking services state behavioral guidelines, users feel free to express themselves as if no restrictions exist (Cemal, 2014). Although women have reaped significant benefits from the internet and social media in terms of power and visibility, as well as access and opportunity, they are also vulnerable to hate speech in the digital realm (Scheuber, 2012). Women are specifically targeted by online hate speech, according to research, and this damages women in long-term ways. It violates women’s fundamental rights and freedoms, as well as their dignity and equality, and has a negative influence on their life on all levels. It has an effect on their mental health and well-being, as well as their social and financial development, resulting in societal costs (European Union, 2014). Twitter, a social networking platform, was completely unknown just a few years ago. However, it soon became very popular among the wealthy and famous, resulting in a tremendous increase in users from the general public
3 B) HATE SPEECH ON SOCIAL MEDIA PLATFORMS

Twitter is a social media service that allows users to send realtime messages known 
as tweets. Tweets are limited to 280 characters and can contain photographs, videos, 
references, and resources (Karle, 2020). Users on Twitter can follow each other. You 
can see someone’s tweets in your Twitter ‘timeline’ 
if you follow them. You 
can either write your own tweets or retweet information that has already been tweeted. 
Retweeting allows for the rapid and efficient dissemination of information to a large 
number of people. Users on Twitter can follow each other. You can see someone’s tweets in your 
Twitter ‘timeline’ if you follow them. You 
can either write your own tweets or retweet information that has already been tweeted. Retweeting allows for the 
rapid and efficient dissemination of information to a large number of people (help.twitter.com). A hashtag is a 
phrase preceded by the hash sign (#) that is used on social networking websites and apps, particularly Twitter. 
Topics are frequently marked with hashtags. This is mostly done to make their tweets more visible (Agarwal 
et al., 2011). These witty phrases originated on Twitter and have since spread to practically every social media 
platform.

2 a) Hate speech in the Middle East

One of the bad consequences of the growth of social media platforms has been the proliferation of hateful 
comments posted on these platforms. Social media has created an environment in which many people consider 
 it appropriate to make derogatory and threatening comments without remorse (Hayes, 2014).

The use of the Arabic language in social media is widespread and constantly increasing. The Arab Social 
Media Report estimates that as of 2017, Facebook users in the Arab region accounted for 8.4% of all Facebook 
users, with more than 150 million Arab users. The total number of monthly active Twitter users in the Arab 
region is estimated to reach 11.1 million in March 2017, making up about 4% of all Twitter users (Salem, 2017).

According to the SkyLine International report on monitoring hate speech in the Middle East, there was an 
increase in incitement and hate speech in the media sectors in 2019, which leads to violence, chaos, conflict and 

Women in the Middle East face increasing harassment and online hate speech as they continue to fight sexual 
harassment on the streets. Over the years, the digital space has often proven to be a hostile environment for 
women around the world, as many have engaged in hacking, breaches of privacy, online smear campaigns that 
threaten freedom of expression, and increasing gender-based violence (Leslie, 2020). According to a 2019 report 
by the Arab Center for the Advancement of Social Media, one-third of young women surveyed said they had been 
exposed to violence and online hate speech, including having their accounts hacked, their personal information 
posted and inappropriate pictures posted.

Hate speech has been spreading recently, especially in the Middle East (Billingham & Bonotti 2019). During the 
Arab Spring, people protested for their rights regardless of their gender (Bligen, 2019). However, it is well known 
that Middle Eastern societies are conservative. Therefore, women’s participation in such actions is generally 
not accepted by the public; for example, causing hate speech against women participating in demonstrations via 
social media (Andersen & de Silva, 2018). Due to the increasing hate speech on the Internet, this study aims to 
examine the role played by social media in the spread of hate speech against women in the Middle East.

3 b) Hate Speech on social media platforms

Online social media platforms and microblogging websites attract internet users more than any other website. 
The services offered by Twitter, Facebook and Instagram are becoming more and more popular among people of 
different backgrounds, cultures and interests. Their content is growing rapidly and is a very interesting example 
of so-called big data. Big data, automatic analysis of people’s opinions and structure/distribution of users in 
 networks etc. It attracts the attention of the researcher who is interested in (Watanabe et al., 2018).

While these platforms offer an open space for people to discuss and share their thoughts and ideas, their nature 
and the large number of posts, comments and messages exchanged make it nearly impossible to control their 
content. Moreover, given the different backgrounds, cultures and beliefs, many people tend to use offensive and 
hateful language when arguing with people who do not have the same background (King and Sutton, 2013).

The increasing popularity of social media platforms such as Twitter for both personal and political 
communication has led to an increase in the number of users on these platforms (Stieglietz & Dang-Xuan, 2013). 
Twitter is the most popular social media tool among internet users worldwide with 500 tweets per day (Mallek 
et al., 2017).

Twitter is a social media tool where users can send very short messages known as tweets. Tweets are short 
messages that are limited to 140 characters in length, but the length has recently changed and the number of 
characters has been expanded to 280 (Hansen et al., 2019). And these (quick and text messages) and are also 
used by users to convey their messages briefly, which makes Twitter different from other social media platforms.

In recent years, social media platforms (especially Twitter) have been used to spread hate messages. According 
to studies by Singh and Diamond (2020: 140), there are approximately 500 million tweets, racial insults or hate 
speech on Twitter per day (Felmlee et al., 2019). Undoubtedly, social media has become a place where people 
can express their anger and hatred without punishment.

For example, the total number of Twitter users in Arab countries is currently more than 11 million with 27.4 
million tweets per day (Alruily, 2018). Hate speech has become a phenomenon in Arab social media. Online
hate speech can facilitate the banning of toxic text content. The complexity, non-formality, and ambiguity of the Arabic dialects prevented the provision of necessary resources for Arabic hate speech detection research (Mulki et al., 2019).

The volatile political/social atmosphere in Middle Eastern countries has always been associated with intense debate; Much of it took place on Twitter. With the participation of more than one opposing party in such discussions, related tweets contain hate speech [Salem, 2017]. Twitter in the Middle East has changed since the Arab Spring. Activists are given the opportunity to spread their message by reaching an audience they could only dream of before the internet. It has been revealed that the number of online hate speech in Middle Eastern countries has increased since the Arab Spring (Murthy, 2018). The increase in hate speech on Twitter may be due to women taking to the streets to demand their rights. As it is known, since the geography of the Middle East is conservative, such actions of women on the street are not accepted by the public. Women’s participation in protests led to hate speech against them through social media.

Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language Processing (Kharde and Sonawane, 2016). Avinash et al. (2017) proposed a novel metaheuristic method (CSK) which is based on K-means and cuckoo search for twitter sentiment analysis. The proposed method has been used to find the optimum cluster-heads from the sentimental contents of Twitter dataset. The efficacy of proposed method has been tested on different Twitter datasets and compared with particle swarm optimization, differential evolution, cuckoo search, improved cuckoo search, gauss-based cuckoo search, and two n-grams methods. Experimental results and statistical analysis validate that the proposed method outperforms the existing methods. Concavar (2013) investigated how hate speech finds a place in the new media and how this discourse is put into circulation as a result of the features of the new media. The study revealed that the relationship between hate speech and the mechanisms of power and the media is a result of the structure of language and ideology. Ring (2013) made a recommendation to encourage self-regulation on the part of social media companies, which involves a move from a “.com” generic top-level domain to one called ”.social.” In order to be part of the consortium of companies included on the “.social” domain, which will hopefully include YouTube, Facebook, Twitter, Instagram and others, an organization must abide by the industry-developed, uniform rules regarding what kinds of hate speech content are and are not permitted on these sites.

4 II.

5 Study Area

The study area is located in the Middle East region. The Middle East is divided geographically into three regions: For this study, Jordan, Saudi Arabia, Lebanon, Qatar, Oman, United Arab Emarates, Kuwait, Bahrain and Egypt were selected. Countries of North afraica were excluded as they have their own accent and terms of hate speech, while Syria, Iraq, Palestine and Y amen were excluded due to the unstable political situation in it. Figure 1 shows the location of the study area. In addition to that, number of females, number of media users and number of Twitter users were collected to be used in the analysis.

6 Methodology

Hate speech mapping started with data collection by designing and distributing a survey to women in the study area. The questionnaire was designed to capture data through an online Survey. A total number of 3850 responses were collected, only 3600 completed responses were included in the analysis. The survey consisted of two parts; one asking about the most hate speech words used in the area and the second part to rank the words based on their negative impact on women. Table 2 shows the results of the survey in which 5 is the most negative impact hate speech word and 1 is the least. Sentiment analysis was applied on the collected words for the last three years using Python program designed for this purpose, a polarity of 0.2 is used in this analysis: if the polarity is greater than 0.2, then the sentiment is positive, while if the polarity is less than -0.2 then it considered as negative, and zero polarity means that the sentiment is not clear. The collected numbers of the Tweets then used in overlay analysis using GIS system. Table 3 shows the number of tweets of each hate speech word in each country as obtained from the sentiment analysis. GIS Spatial Analyst tools applied overlay analysis on the collected numbers of the Tweets. Overlay analysis is a group of methods applied in optimal site selection or suitability modeling. It is a technique that applies a common scale of values to diverse and dissimilar inputs to create an integrated analysis (Al- Omari et al, 2020). Overlay analysis often requires the analysis of many different factors that may not be equally important. Even within a single raster, one must prioritize values. Prioritization values have been done in weighted overlay method based on the opinions of the respondents. Figure 2 shows the followed methodology in this study.

7 Results and Discussion

The distribution of the hate speech based on the negative impact of each word with respect to the total number of tweets in each country is shown in Figure ?? It is not necessary to classify the hate speech in the countries
Based on this factor only, as the number of tweets is relative to the total number of media users and Twitter users as well.

Volume XXII Issue X Version I 36 ( ) Based on Figure ??, the results showed that Egypt is the most country in the study area of negative impact hate speech words followed by Saudi Arabia. Figure 5 illustrates the results from using overlay method in GIS Spatial Analysis. In this figure the study area is divided into 5 categories based on the hate speech. Very high hate speech area, high hate speech area, moderate hate speech area, low hate speech area and very low hate speech area as summarized in table 4. The results showed different distribution of hate speech according to the all factors. Saudi Arabia and Egypt came in the very high class of hate speech. Lebanon is considered a high hate speech area, while Jordan and United Arab Emirate are moderate hate speech areas. In contrast, Oman and Kuwait are considered as low hate speech areas, and in the last place Qatar and Bahrain.

8 V. Conclusions and Recommendations

In this research, the primary focus is on mapping hate speech in the study region. The suggested technique also attempted to categorize the region into hate speech classes based on a variety of characteristics ranging from very high to extremely low levels of hate speech. The data for the overlay approach that was employed in this research came from replies to an online survey that was issued to the women in the study’s geographic region. From the responders, the words with the greatest negative effect were gathered and passed into Python code that assessed hate speech terms over a period of three years in the past. The findings revealed that it is not essential that Saudi Arabia and Egypt be both classified as having extremely high hate speech levels, but Jordan and the United Arab Emirates are classified as having moderate hate speech levels. According to the findings, Qatar and Bahrain are in the extremely low hate speech classification category.

Volume XXII Issue X Version I 40 ( )

Figure 1: Study Area in the Middle East region

© 2022 Global Journals
Figure 2: Year 2022 A
Figure 3: Figure 2:
Figure 4: byA
Figure 5: Figure 3 :Figure 4 :

Figure 6: Table 1 :

<table>
<thead>
<tr>
<th>Country</th>
<th>Population (Millions)</th>
<th>No. of Females</th>
<th>No. of media users</th>
<th>No. of Twitter users</th>
</tr>
</thead>
<tbody>
<tr>
<td>United Arab Emirates</td>
<td>9.94</td>
<td>3081400</td>
<td>9840000</td>
<td>2106000</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>35.08</td>
<td>14803760</td>
<td>27800000</td>
<td>7968000</td>
</tr>
<tr>
<td>Qatar</td>
<td>2.91</td>
<td>724590</td>
<td>2870000</td>
<td>36252</td>
</tr>
<tr>
<td>Oman</td>
<td>5.16</td>
<td>1754400</td>
<td>4140000</td>
<td>33032</td>
</tr>
<tr>
<td>Lebanon</td>
<td>6.80</td>
<td>3372800</td>
<td>4370000</td>
<td>49093</td>
</tr>
<tr>
<td>Kuwait</td>
<td>4.30</td>
<td>1664100</td>
<td>4250000</td>
<td>87500</td>
</tr>
<tr>
<td>Jordan</td>
<td>10.24</td>
<td>5058560</td>
<td>6300000</td>
<td>37149</td>
</tr>
<tr>
<td>Egypt</td>
<td>103.3</td>
<td>51133500</td>
<td>49000000</td>
<td>2960000</td>
</tr>
<tr>
<td>Bahrain</td>
<td>1.50</td>
<td>605440</td>
<td>1500000</td>
<td>23652</td>
</tr>
</tbody>
</table>

III.
2

<table>
<thead>
<tr>
<th>Hate Speech Word</th>
<th>Spinster</th>
<th>Unmarried</th>
<th>Your place is in the Kitchen</th>
<th>You're a</th>
<th>Women</th>
<th>Have Half a Brain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most Negative impact</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Figure 7: Table 2:

3

<table>
<thead>
<tr>
<th>Country</th>
<th>Spinster you're a girl</th>
<th>Unmarried</th>
<th>your place is in the kitchen</th>
<th>Women have half a brain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bahrain</td>
<td>1406</td>
<td>1005</td>
<td>403</td>
<td>545</td>
</tr>
<tr>
<td>Egypt</td>
<td>6647</td>
<td>6704</td>
<td>3716</td>
<td>2999</td>
</tr>
<tr>
<td>Jordan</td>
<td>3078</td>
<td>3433</td>
<td>1074</td>
<td>851</td>
</tr>
<tr>
<td>Kuwait</td>
<td>3292</td>
<td>2747</td>
<td>2203</td>
<td>1659</td>
</tr>
<tr>
<td>Lebanon</td>
<td>2598</td>
<td>1840</td>
<td>1959</td>
<td>1559</td>
</tr>
<tr>
<td>Oman</td>
<td>2373</td>
<td>1875</td>
<td>1574</td>
<td>1101</td>
</tr>
<tr>
<td>Qatar</td>
<td>639</td>
<td>748</td>
<td>683</td>
<td>309</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>5056</td>
<td>3169</td>
<td>1469</td>
<td>1079</td>
</tr>
<tr>
<td>United Arab Emirates</td>
<td>2474</td>
<td>418</td>
<td>326</td>
<td>735</td>
</tr>
</tbody>
</table>

Figure 8: Table 3:

4

<table>
<thead>
<tr>
<th>Hate Speech Class</th>
<th>Very high</th>
<th>High</th>
<th>Moderate</th>
<th>Low</th>
<th>Very low</th>
</tr>
</thead>
</table>

[Note: Country Saudi Arabia Egypt Lebanon Jordan, United Arab Emirate Oman, Kuwait Qatar, Bahrain]

Figure 9: Table 4
Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.


V. CONCLUSIONS AND RECOMMENDATIONS


[Skyline (5)] Skyline. Monitors Hate Speech in Arab Media Raporu. (Stokholm) 2019. (Skyline International)


[Salem (7)] Social Media and the Internet of Things towards Data-Driven Policymaking in the Arab World, F Salem. 2017. MBR School of Government. Dubai


