Hate Speech Patterns in Social Media: A Methodological Framework and Fat Stigma Investigation Incorporating Sentiment Analysis, Topic Modelling and Discourse Analysis

Vajisha Udayangi Wanniarachchi
School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand
V.U.Wanniarachchi@massey.ac.nz

Chris Scogings
School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand

Teo Susnjak
School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand

Anuradha Mathrani
School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand

Abstract

Social media offers users an online platform to freely express themselves; however, when users post opinionated and offensive comments that target certain individuals or communities, this could instigate animosity towards them. Widespread condemnation of obesity (fatness) has led to much fat stigmatizing content being posted online. A methodological framework that uses a novel mixed-method approach for unearthing hate speech patterns from large text-based corpora gathered from social media is proposed. We explain the use of computer-mediated quantitative methods comprising natural language processing techniques such as sentiment analysis, emotion analysis and topic modelling, along with qualitative discourse analysis. Next, we have applied the framework to a corpus of texts on gendered and weight-based data that have been extracted from Twitter and Reddit. This assisted in the detection of different emotions being expressed, the composition of word frequency patterns and the broader fat-based themes underpinning the hateful content posted online. The framework has provided a synthesis of quantitative and qualitative methods that draw on social science and data mining techniques to build real-world knowledge in hate speech detection. Current information systems research is limited in its use of mixed analytic approaches for studying hate speech in social media. Our study therefore contributes to future research by establishing a roadmap for conducting mixed-method analyses for better comprehension and understanding of hate speech patterns.

Keywords: social media, hate speech, sentiment analysis, topic modelling, discourse analysis, fat stigma.

1 Introduction

Social media has become an integral part of everyday lives enabling users to freely voice their opinions anytime over online platforms. The number of social media users is seen to be growing at a rate of 13% on a year-on-year basis, with current over 4.20 billion users, which represents 53% of the world's total population (Kemp, 2021). Online platforms have changed the way we interact with each other, and the way we acquire news (Gallacher, Heerdink, & Hewstone, 2021), enabling users to freely express themselves using emotive language when arguing over contested topics. Unfortunately, users may target certain community groups in these emotional exchanges which could promote negativity or discrimination against them (Vigna et al., 2017). Communities can be aligned to multiple identities including that of race, religion, gender or sexual orientation, which can in turn lead to multiple levels of discrimination (Mossie & Wang, 2020). Behavioural researchers are therefore keen to make sense of how public opinions are framed around certain topics or what prejudicial views are expressed in online conversations. Offensive behaviours are often motivated by the anonymity facilitated by social media platforms as people are known to be more comfortable in expressing provocative views when they know they are anonymous (Erjavec & Kovačič, 2012). While anonymity has a liberating impact on users, it also reduces their hesitation when expressing contempt toward those posts that they do not agree with (Tucey, 2010). Anonymity can lead to 'deindividuation and disinhibition' when users let go of their emotions without fear of being judged for their aggressive posts while discussing controversial topics (De Brún et al., 2014, p. 74). It liberates users from being held accountable for their hateful messages (Delgado & Stefancic, 2014) as they openly reveal their prejudices (Mondal, Silva, & Benevenuto, 2017). Hate speech can be thus advanced and cause further divisions based on religion, colour, race and body type amongst others.

Users often experience more insults in the online space compared to physical spaces, as the Internet is not designed to safeguard users from online abuse (Zhong, 2020). Therefore, social media platforms can promote stigmatization if users feel safe in posting negative content online. It is indeed challenging to mitigate the impact of hateful commenting without having a proper understanding of the types and targets of hate speech, as comments contain overlapping targets (Salminen et al., 2018). Hateful comments may target various communities based on their sexual orientation, ethnic origin, religious faith, gender identity or body weight or any combination of these (Albadi, Kurdi, & Mishra, 2018; Matamoros-Fernández & Farkas, 2021; Ștefăniță & Buf, 2021; Wanniarachchi et al., 2019). The offensive content posted online is therefore very subjective with no clear definition of the target (i.e., the person or community being addressed in the hate speech); hence, human interpretations of societal events must complement the data-centric computational approaches for analysing its hateful (Kocoń et al., 2021). Among the many forms of hate speech, weight-based teasing or fat shaming or anti-fat talk reinforces the 'cultural preoccupation of obesity as individual responsibility' to further perpetuate weight discrimination (Brewis, SturtzSreetharan, & Wutich, 2018, p. 1). As such, body weight is perceived with a judgemental stance, with obese individuals having to experience fat stigma on an ongoing basis.

The past three decades have witnessed a startling increase in worldwide obesity rates (WHO, 2022); moreover, obesity has been linked to metabolic disorders, cardiovascular diseases and other chronic illnesses (Rawal et al., 2020; Szepietowska et al., 2016). Alongside health risks, other perceptions associated with 'ideal' body shapes and sizes too have evolved. Body image is described as cognitions, perceptions, and attitudes towards one's appearance, where an individual may measure their bodily features with some overall level of attractiveness (Izydorczyk et al., 2021). When such forms of 'ideal' body image perceptions become prevalent in mass media and social media, they can reinforce thinness as a form of ideal body size, leading to more expectations across body shapes in society (Maftei & Merlici, 2022; McComb & Mills, 2022; Veldhuis, Konijn, & Seidell, 2014). As such those people who do not fit within the ideal body classification face much discrimination. The social devaluation and denigration by blaming and shaming individuals based on their weight gain lead to fat stigmatization (Tomiyama, 2014). This could be in many forms, such as derogatory humour, sarcastic comments or unflattering visuals amongst others. Consequently, these individuals experience feelings related to body dissatisfaction and anxiety, which can further result in eating disorders, depression or poor self-esteem (Albadi et al., 2018; Larson, 2021; Murakami & Latner, 2015). A theorised pathway linking gender with body image by Fredrickson and Roberts (1997) is of the view that women tend to take an observer's view as the primary view of their bodies. This perception can lead women to continuous self-monitor their bodies which could increase self-shaming behaviours. Fat studies have revealed higher rates of internalized stigma and risks of weight-based discrimination toward women in education and employment (Himmelstein, Puhl, & Quinn, 2017; Puhl, Andreyeva, & Brownell, 2008), and that, obese female children/adolescents encountered more teasing and social marginalization (Almenara & Ježek, 2015; Tang-Péronard & Heitmann, 2008).

Hate speech expressions over social media, specifically hate content targeting excessive weight (or fatness) have been much studied (i.e., Brooker et al., 2018; Chou, Prestin, & Kunath, 2014; Holmberg et al., 2018; Hussin, Frazier, & Thompson, 2011; Jeon et al., 2018; Kent et al., 2016; Lydecker et al., 2016); however, very few of these have used natural language processing (NLP) techniques (Wanniarachchi et al., 2022). Prior studies are found to be limited to manual analysis techniques that make use of small text-based datasets acquired from social media, but recent advances in machine learning allow for large datasets to be effectively analysed computationally. For instance, sentiment analysis enables text mining of large amounts of user opinions (in some given context) to reveal overall attitudes and the emotional content of these expressions (Liu, 2012). It can facilitate the detection of hate content implanted in open-ended discussions by extracting the negative sentiments associated with sensitive topics. Further, topic modelling, a machine learning technique, can be used to draw out common themes from the unstructured textual data within a given linguistic corpus (Wallach, 2006). In this manner, a systematic assessment of topics, sentiments and emotions embedded in social media conversations can be undertaken to reveal hate speech patterns (Lipizzi, Iandoli, & Marquez, 2015). Additionally, discourse analysis can augment these computer-mediated quantitative approaches with a qualitative perspective. Discourse analysis is a linguistic methodology that identifies social ideology on various topics through language analysis. It has been used to identify societal stances on topics such as racism, feminism or sexism (Chiril et al., 2020; Durrheim et al., 2018; Thompson, Rickett, & Day, 2018). Therefore, a mixed-method analytical approach comprising machine learning techniques (e.g., sentiment analysis, emotion analysis, topic modelling) and discourse analysis, can help overcome current limitations in detecting hate speech (fat stigma) patterns (Chou et al., 2014; Yeruva, Junaid, & Lee, 2019) to offer richer and more meaningful conversational insights.

2 Research objectives

Prior studies have asserted that researchers must expand computational (or quantitative) capabilities with societal (or qualitative) interpretations when investigating any media-facilitated public discussions (Brooker et al., 2018; Jeon et al., 2018; Lydecker et al., 2016) (Brooker et al., 2018; Jeon et al., 2018; Lydecker et al., 2016). That is, while computational approaches detect linguistic patterns within the data, qualitative approaches render meaning to the patterns that have been produced (Ophir, Walter, & Marchant, 2020). The computational patterns that are extracted from empirical data need descriptive explanations to glean insight and convincingly represent its meaning for informing theory and practice (Miranda et al., 2022). Following these calls from communication researchers, our study proposes a combination of analytical approaches for examining online discourses that are framed around hate speech (containing fat stigmatizing content).

Two objectives are laid out in this investigation. First, we outline a novel mixed-methods framework that provides a synthesis of quantitative and qualitative text analysis approaches for detecting hate speech patterns from a large linguistic corpus. Multiple methodological lines of inquiry can provide more purposeful reflexivity in the framing of narratives within their given context to generate new insights (Ophir et al., 2020). To that end, our framework establishes a roadmap for applying multiple approaches like sentiment analysis, topic modelling and discourse analysis to detect hate patterns in social media discussions. Second, we demonstrate how this framework informs the data analysis and interpretation of results by using large unstructured textual data extracts from two popular social media platforms. The context of our investigation is a practical hate speech example that specifically relates to discrimination of fat individuals. Different data-driven analysis techniques are applied to detect underlying fat stigmatizing patterns embedded in social media conversations. Current information systems research is much limited in their use of mixed analytic approaches to interpret patterns within a given corpus of texts. Information systems researchers need to consider emergent computational tools in conjunction with traditional qualitative (and quantitative) approaches to construct empirical patterns across large populations and inform theory (Miranda et al., 2022). This study therefore contributes to methodological design by combining multiple analytic frames that together enable meaningful interpretation of large amounts of textual data for identifying linguistic patterns within a data corpus.

3 Literature review

Hate speech comprises public statements made in discriminatory language intended to denigrate specific individuals or community groups (Delgado & Stefancic, 1991). This section elaborates on prior literature to highlight methodologies that have been used to detect hate speech and fat stigma in social media. We specifically focus on the application of data-driven analysis techniques. Our findings indicate sentiment analysis and topic modelling are key techniques currently used for hate speech detection from unstructured textual social media data extracts, followed by emotion analysis and discourse analysis. Next, we delve into how these detection techniques have been applied for identifying fat stigma patterns. Different literature streams have helped lay out a methodological framework (in section 4) for conducting a robust analysis that is substantiated with a logical chain of evidence when combining NLP techniques with other pattern matching methods (e.g., word frequencies, discourse analysis).

3.1 Detecting hate content in social media

Many studies on hate speech detection in social media have been inspired by sentiment analysis and topic modelling methods. While some studies have used sentiment analysis, topic modelling, discourse analysis or emotion analysis, few have suggested a combination of these.

Emotional profiles of social media comments were analysed by Martins et al. (2018) to classify hate speech based on different emotions and their intensity. Their study highlights that emotion analysis can aid in improving the accuracy of hate speech detection. Emotion analysis, along with sentiment analysis was also used by Rodríguez, Argueta, and Chen (2019) to cluster relevant Facebook pages and filter negative posts/comments that promote hate speech. Their study indicates the importance of studying different aspects of hate speech by identifying sentiments and emotions of hate speech and integrating k-means clustering to determine the most frequently discussed topics. Sentiment analysis was also employed by Irani et al. (2020) to study social media data in the Indonesian language for the study of radicalism in Indonesia. They have used word embeddings (word2vec) to detect the top-most radical words and combined Restricted Boltzmann Machine and Back-Propagation (RBMBP) network to enhance the classification of hate speech content which leads to radicalism. al-Utbi (2019) underlines the usage of Critical Discourse Analysis (CDA) to determine hate speech in social media. The author studied Facebook posts targeting Islam and Muslims; they found the language of the posts and the semiotic details to indicate the existence of expressed hatred.

When identifying commonly discussed topics in hate speeches, topic modelling has been used to reveal rich information about user preferences, emotions and associated behaviours. Topic modelling and deep learning were used by Alshalan et al. (2020) to study hate speech tweets during the COVID-19 pandemic. Using Convolutional Neural Network (CNN) and Non-Negative Matrix Factorization (NMF), they detected the most hateful tweets and identified the main themes. Overall, the number of non-hate tweets were found to be higher than hate tweets. Unsupervised topic modelling too has applied to hate speech against immigrants in Spain with the appearance of far-right party Vox by Calderón, de la Vega, and Herrero (2020). Their study discovered underlying themes associated with immigrants in Spain that were similar to the discourse of Vox. Latent Dirichlet Allocation (LDA), an unsupervised topic modelling technique, was used by Mathew et al. (2020) for performing temporal analysis on hate speeches made on posts in gab.com (i.e., a social media site with a lax moderation policy compared to mainstream social media sites like Twitter and Facebook). Using data from different time points, the authors explored various topics of interest and found that hateful users tend to speak about Blacks, Muslims, Jews and politics, while the non-hateful users generally speak about politics, technology, sports and free speech matters.

Furthermore, some researchers have considered both sentiment analysis and topic modelling methods for detecting hate speech and analysing the associated patterns. Deephate, proposed by Cao, Lee, and Hoang (2020), is a novel deep learning model which combines word embeddings, sentiments and topical information to detect hate speech in online social platforms. Authors evaluated their model using three publicly available datasets and demonstrated that Deephate 'outperformed the state-of-the-art baselines' (p. 18) and concluded that a combination of multiple text analysis methods results in effective hate speech detection. Aspect-Based Sentiment Analysis (ABSA) was used by Pronoza et al. (2021) for instance-based hate speech detection in Russian social media. They further observed that the models could benefit from a combination of linguistic and sentiment features with BERT pretraining and an additional dense layer. Shibly, Sharma, and Naleer (2021) used sentiment analysis to validate the sentiment values of a hate speech dataset and combined it with topic classification to identify the most common topics within the dataset. The study observed that the most common hate speech types were related to race, ethnicity, sex and religion. Therefore, the proliferation of computational approaches to detect bias or discriminatory messages from unstructured and unlabelled social media data is growing fast.

3.2 Detecting obesity content in social media

Methods associated with fat stigma studies on social media discourses were systematically reviewed by Wanniarachchi et al. (2022). The review ascertained that textual data analysis in most studies was done mainly using manual qualitative coding approaches, although, few studies have used sentiment analysis, topic modelling and emotion analysis too. Sentiment analysis and word co-occurrence analysis has been used by Wanniarachchi et al. (2019) to reveal gender-based fat stigmatization content in Twitter and YouTube. The findings from this study observed that female body objectification is more commonly displayed in fat stigma content than male body objectification.

Sentiments expressed toward the Fat Acceptance (FA) movement and links between the movement and user characteristics have been studied by Bograd, Chen, and Kavuluru (2022) using sentiment analysis. Their model indicates extreme anti-FA discourse as most FA tweets displayed 'opposing' views. Kent et al. (2016) examined the overall sentiments towards the topics of cancer and obesity using sentiment analysis and employed manual coding to uncover discussion themes that connect cancer and obesity. Sentiments indicated the "dominance of negative framing around obesity" (p. 457). Lazarus et al. (2021) examined stigma associated with Non-Alcoholic Fatty Liver Disease (NAFLD), Non-Alcoholic SteatoHepatitis (NASH) and obesity using Twitter data. The study analysed sentiments, geographic distribution and hashtag patterns of obesity-related content and detected much negativity in the majority of tweets on obesity. Chou et al. (2014) also employed sentiment analysis and qualitative analysis to study social media discussions related to obesity. The study confirmed the stigmatization of fat people and revealed patterns of such discussions, such as themes and sentiments. The combination of multiple text analysis methods was also tested by Yeruva et al. (2019) to examine the relationship between obesity and healthy eating. Sentiment analysis, co-occurrence analysis and topic modelling aided the authors in understanding the social media influences on behaviour and decision-making for healthy eating and obesity prevention.

While recent studies conducted for detecting hate speech and fat stigma in social media have used sentiment analysis and/or topic modelling to provide evidence that such techniques can capture hate speech in social media (i.e., Bograd et al., 2022; Chou et al., 2014; Kent et al., 2016; Yeruva et al., 2019), these are limited in their application. The literature review reveals very few studies have analysed hate speech discourses surrounding fat stigma that can show meaningful semantic patterns. Uncovering behavioural aspects such as what perceptions revolve around discussions on exercise/diet plans or unhealthy lifestyles will help create a better understanding of the underlying stigmatizing content. Discourse analysis can enable social scientists to recognize emergent stigma patterns present in the given social context. Moreover, prior studies have mostly considered Twitter as their single data source (Wanniarachchi et al., 2022); but, with the text-limit constraints imposed on tweets, users have to adapt their intended message within the specified data bounds. We propose that by combining data from multiple social media platforms, we can inspect richer content. Reddit is one such contemporary social media platform that facilitates candid naturalistic expressions from its users to open up many methodological possibilities for conducting thematic analysis (Boettcher, 2021). Reddit and Twitter platforms can together assist in unravelling the hidden patterns in hate speech, specifically for fat stigma.

The use of novel text mining and machine learning methods along with in-depth text analysis techniques such as discourse analysis can aid in the discovery of hate speech patterns. However, we find a gap in the provision of guidelines or a frame of reference that encompasses multiple methods for facilitating detection and enables researchers to comprehend the stigma patterns embedded in social big data. The next section builds on these shortcomings wherein a methodological framework is proposed. Steps on data collection, data pre-processing and data analysis for classifying different user perspectives are outlined for enabling researchers in deducing patterns that project a more holistic picture.

4 Methodological framework

Having an overarching methodology lends support to the researcher by providing them an understanding of the data context and establishing proper checks for collecting empirical data that align with the study's purpose. It further ensures that appropriate analytical approaches are subscribed for examining and interpreting the data.

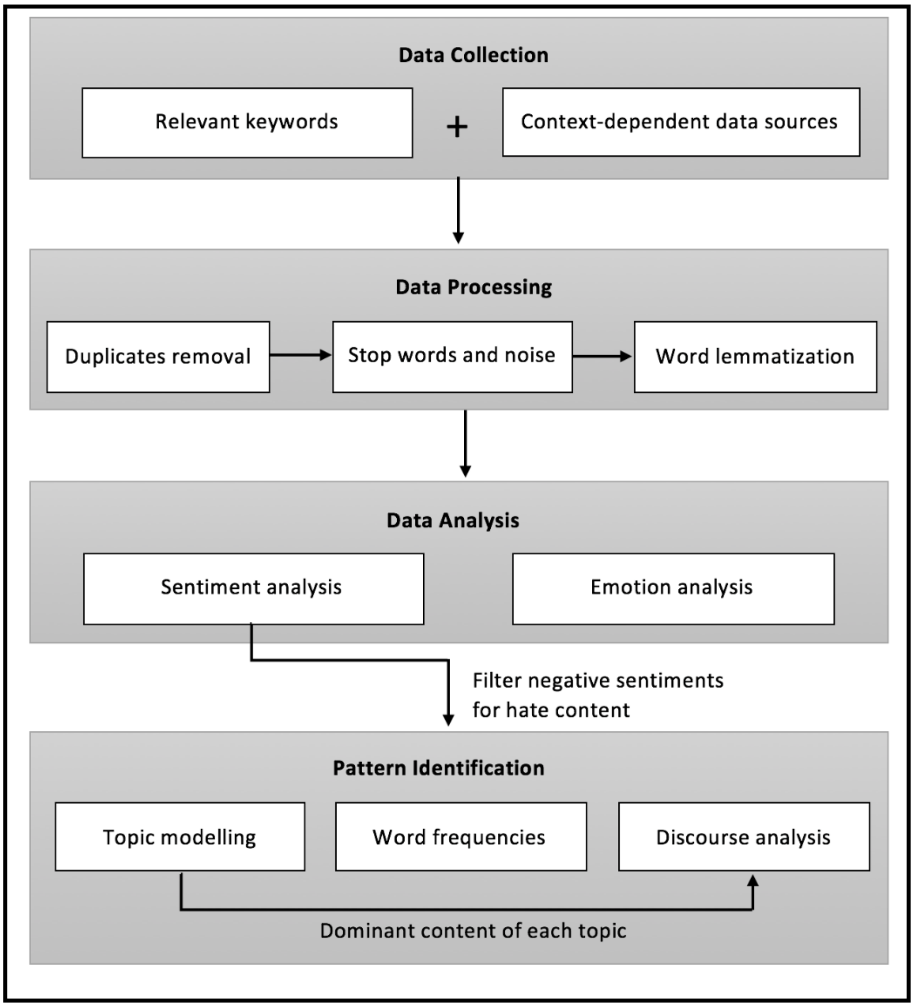


Figure 1. Proposed methodological framework

Moreover, a graphical illustration aids in understanding concurrent and sequential steps as researchers move back and forth with their data analysis. Having a high-level framework that identifies broad activities needed within each methodological step provides operational flow and precludes potential mistakes of missing out some tasks within each step. The proposed methodological framework, outlined in Figure 1, provides a detailed representation of what tasks are required at each of the identified steps, namely data collection, data pre-processing, data analysis and pattern identification.

This study aims to uncover underlying patterns within textual content extracted from social media to identify fat stigmatizing patterns. Moreover, our proposed framework can be applied to other forms of hate speech as well. It follows NLP workflow for data collection and data pre-processing and subsequently combines sentiment analysis, emotional analysis, topic modelling, word frequencies and discourse analysis to assess fat stigma patterns in different social media platforms. The outputs of sentiment analysis can aid in filtering the discourse, for instance, only negative sentiments and the associated negative content can be used as inputs for conducting topic modelling to identify fat stigmatizing themes. Apart from discovering these topics/themes, topic modelling can highlight the most dominant posts/comments within each topic, which can then be used to carry out discourse analysis. This framework is based on the research directions implied by recent literature and is further validated by the authors' experiences in analysing fat stigma content in social media. The following subsections describe the activities pertaining to each step.

4.1 Data collection

Data collection starts with a prior understanding of the context, that includes the selection of relevant keywords and determining which data sources need to be explored. Therefore, relevant literature must be examined to identify pertinent keywords that reflect the chosen context and serve as a starting point for collecting data. For example, this study focuses on uncovering fat stigma, accordingly, keywords have been determined from obesity-related literature. Keywords comprising 'fat', 'overweight' and 'obesity' were thus identified (i.e., Chou et al., 2014; So et al., 2016) and these informed the text mining process. Moreover, based on the direction of the research path, the keywords can be built upon further. That is, if a study seeks to investigate gender aspects in fat stigma discussions, other keywords associated with gender are added to the keyword list (i.e., 'fat+girl', 'obese+man', etc.).

Next, relevant social media data sources need to be determined. Empirical data that is context-dependent on the selected keywords are subsequently scraped for further analysis. Fat studies have made significant use of Twitter in comparison to other social media platforms (Wanniarachchi et al., 2022); however, other platforms, such as Reddit (which appears "more as a discussion forum, and [its] comments have a wider range of document lengths than Twitter data" (Curiskis et al., 2020)) too are gathering momentum. Twitter and Reddit platforms served as our primary data sources. Further, researchers must scope their study to the type and format of digital data that must be used in the knowledge discovery process. This study's context refers to the discovery of linguistic patterns concerning fat-stigma; hence, corpora of texts extracted from obesity-related discussions informed our investigation. The publicly available data scraping tools or APIs (Reddit, 2020; Twitter, 2019) assisted in collecting textual data for inclusion in the corpus. Since a corpus of texts is characterised by words that are independent of any non-text format (i.e., image, video, etc.), conducting sentiment analysis on the observed texts is rather straightforward (Curiskis et al., 2020; Wanniarachchi et al., 2022).

A major issue often encountered while collecting social media data is of noise or the acquisition of irrelevant data. For instance, one example of irrelevant data relates to acquiring text content posted by a user whose username matches the keywords used for scraping data. If a keyword combination is ‘fat + girl’, then comments posted by a user named ‘fatgirl123’ would be scraped even though they are not relevant to the topic under consideration. Or, in another case, while the given keyword exists in the scraped comment, the context of its use is different. That is, the keyword ‘fat’ could be used in different scenarios, such as, the comment has a line containing words like ‘big fat liar’. Such cases can be partially solved by using a combination of keywords relevant to the study's context instead of just one keyword; although when the length of the textual content is large, this method may not be workable. In our case, we found that a combination of keywords provided us with more relevant data from Twitter compared to Reddit, as Twitter enforces a text-limit on each post.

4.2 Data pre-processing

Pre-processing steps assist in presenting the data in a more structured format for conducting effective and accurate analysis. Pre-processing techniques, including removal of duplicates, noise and stop words and word lemmatization, are described next.

4.2.1 Duplicates removal

When scraping data from social media, one of the key issues is data duplication. On Twitter, most of the duplicates occur with re-tweets. Although re-tweets can be eliminated at data collection stage for Twitter, the duplicated posts on Reddit must be explicitly removed. Duplicated posts occur on Reddit because users often post the same content on multiple subreddits. The ‘remove duplicates’ feature in Microsoft Excel can aid in resolving this (Microsoft Corporation, 2016).

4.2.2 Stop words and noise removal

Creating insights from raw textual data is complicated; moreover, unnecessary data can cause distractions to produce results that are inconsistent in the subsequent downstream tasks. Therefore, removing stop words and noisy data are important pre-processing tasks. Common stop words include prepositions (of, to, for) and conjunctions (and, or also) make a language functional rather than carry information (Boban, Doko, & Gotovac, 2020). Additionally, other superfluous characters such as symbols (@, !, *, etc.), emojis (:D, :’(, etc.) and URLs can add complexity, that makes the overall analysis process harder to execute. Hence, all such noisy characters need to be removed.

4.2.3 Word lemmatization

Stemming and lemmatization are used to convert a word to its basic form. Stemming refers to a crude heuristic process that truncates words in the hope of achieving their primitive meaning. This often includes the removal of derivational affixes. Lemmatization uses “vocabulary and morphological analysis for removing the inflectional endings and returning the word to its base or dictionary form, also known as the lemma” (Manning, Raghavan, & Schütze, 2008, p. 32). However, with stemming, the resulting dataset may contain some words with no meaning. Lemmatization does not result in meaningless words since token words are presented in their base form by using their origins, irrespective of whether the text is used as a verb or as a noun (refer to Table 1). Therefore, lemmatization plays an important role in text mining applications.

Token word	Stemmed word	Lemmatized word
Troubled	troubl	trouble
Saw	s	see

Table 1. Stemming and lemmatization

4.3 Data analysis

The researcher needs to conduct a meaningful analysis of the pre-processed data for establishing generalizations. The data analysis provides overall sentiments and emotions of the hate speech data gathered and cleansed in the previous two steps. Sentiment analysis can reveal the users' attitudes (which may be positive, negative or neutral) as expressed in the selected hate speech context, while emotion analysis indicates feelings based on the choice of words (e.g., anger, surprise, trust, etc.). Data analysis, in this study's context, is deciphering of the fat stigma content expressed in social media discussions in a cumulative and holistic manner. First, it is important to have a proper grounding of the overall themes around the discussions on fatness. Sentiment analysis and emotion analysis can assist in building themes that are bottom-up and data-driven. The NRC dictionary developed by Mohammad and Turney (2013) can capture in-depth sentiments associated with a linguistic corpus and identify a variety of emotions. Therefore, all of the scraped data (comprising all acquired posts/comments extracted from Twitter and Reddit) are organized into one document for the purpose of executing sentiment analysis. In this manner, we can obtain gross positive and negative sentiment scores from the document with no researcher bias in play. This form of analysis is also beneficial for comparing two sets of data. Next, emotion analysis can be applied to reveal the scores of the eight primary emotions, namely anger, anticipation, disgust, fear, joy, sadness, surprise and trust, for this linguistic corpus objectively. Further, with Pyplutchik (Semeraro, Vilella, & Ruffo, 2021), we can interpret how these emotions are related and highlight which emotions are more prominently expressed.

Subsequently, sentiment analysis using Stanford's coreNLP (Manning et al., 2014), can be performed on specific posts/comments of the corpus to determine their sentiment scores and labels. Sentiment score or valence score is determined by assigning a sentiment value (ranging from -1 for negative to +1 for positive and 0 for neutral) to each word and then by aggregating all sentiment values of the specific post/comment. Therefore, posts/comments with more negative words indicate a higher negative valence score, and posts/comments with more positive words will indicate a higher positive valence. If there are equal amounts of negative and positive words in a text, the valence score would be '0' and be considered a neutral text. Based on these sentiment scores, each post/comment is labelled as a positive, negative or neutral sentiment. Next, by considering those posts and comments that indicate negative sentiment label, we can filter the linguistic corpus to obtain the hateful content. This can then be used as the input for the next stage, that is, to identify hate speech patterns (or fat stigma patterns in this study's context).

4.4 Pattern identification

The purpose of pattern identification is to deduce those prominent features that have a shared societal context (e.g., word frequencies, topics, themes) from the big dataset. Therefore, the researcher must first identify their context and construct meaning based on that context. In this study, our context relates to gaining insights on fat stigma. The content that displayed negative sentiment in the sentence-based sentiment analysis has been used as the input for identifying common topics/themes and for determining those words that are most frequently associated with fat stigmatizing discussions on social media.

Topic modelling based on Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003) can inform on the probability distribution of topics spread over documents, based on their word distribution. Gensim LDA implementation (Řehůřek & Sojka, 2011) has been used to execute topic modelling on filtered fat stigma content. The topic models thus generated were evaluated by first calculating the coherence score. Coherence score is the value of the relative distance between words within a topic. There are two major coherence score types, C_V (which typically range between 0 and 1) and uMass (that range between -14 and 14). This study used the C_V scores as it provides the strongest correlation with human ratings and is considered a very reliable topic coherence evaluation measure (Röder, Both, & Hinneburg, 2015). In order to determine the optimal number of topics within the male and female corpus, we generated multiple models, each with a different number of topics ranging from 2 topics through to 10 topics. We then used the model with the highest coherence score as the most optimal for determining the number of topics for each corpus. Once the optimal number of topics was determined for the two datasets, the models were further tuned and optimized the models using alpha and beta hyperparameters.

Apart from identifying the discussion topics of the dataset, the most dominant topic of each post/comment and the most dominant post/comment for each topic helped determine patterns from our analysis. The most dominant post/comment on each topic has next been used to perform discourse analysis and comprehensively examine the negative social media content. Selecting the dominant content via topic modelling eliminates the subjectivity associated with excerpt selection for discourse analysis and facilitates the examination of the most representative posts/comments in the corpus. Discourse analysis can then be carried out manually by closely examining the text and exploring how social media users use language in relation to the fat stigma context. Concurrently, word frequencies of the data corpus too can be used to identify words that are mostly associated with fat stigma.

5 Fat stigma investigation

This section presents a social media investigation that elaborates how NLP and qualitative methods have together provided multi-level insights from the textual data extracts. We specifically demonstrate the empirical analysis of fat stigma content that was extracted from social media discourses around female and male genders. We believe this empirical evidence will showcase how our proposed methodological framework (shown in Figure 1) can be applied to practical real-world scenarios.

5.1 Data collection and data pre-processing

The analysis has been carried out to understand how males and females are positioned in obesity-related discussions. Therefore, keyword combinations that represented fatness and gender were selected to scrape data (refer to Table 2). As this analysis is focused on textual data only, the publicly available text-based content from Twitter and Reddit was scraped (via their APIs) from 15th October 2021 to 15th January 2021. The data thus collected served as our data corpus.

Obesity-related keywords	**Fat, overweight, obese**
Gender-related keywords	Female, girl, woman, women, male, boy, man, men
Keyword combinations used	fat + (female, girl, woman, women)
	overweight + (female, girl, woman, women)
	obese + (female, girl, woman, women)
	fat + (male, boy, man, men)
	overweight + (male, boy, man, men)
	obese + (male, boy, man, men)

Table 2. Keyword combinations

A closer examination of the number of posts/comments scraped for different keyword combinations were not found to be equally distributed; rather, a previous study found a larger number of posts/comments for the keyword "fat" compared to "overweight" and "obese" (Wanniarachchi et al., 2019). Due to this difference, we selected an equal number of posts/comments representing each keyword combination from each social media platform through random sampling. Data comprising 60000 comments/posts representing every keyword combination across both social media platforms were finally selected. In maintaining equal data distribution, we formed dataset samples consisting of 30000 comments/posts targeting males and an equal number targeting females. Further, each of these datasets (having 30000 comments/posts) in turn comprised 10000 comments/posts representing each keyword combination. From the 10000 comments/posts that originated from Twitter and Reddit, 5000 posts/comments were from Twitter and the other 5000 posts/comments from Reddit. The datasets were cleansed next by removing stop words, symbols and emojis. Later, lemmatization was also applied to further cleanse the dataset for subsequent analysis.

5.2 Data analysis: overall text analysis of fatness

This section discusses the three data analysis approaches that were used to unearth sentiment values, conduct emotion analysis and generate sentiment labels for text-blocks representing posts/comments. The following subsections expand on each of these approaches.

5.2.1 Sentiment analysis

The sentiment value for each selected comment/post has next been calculated using the NRC dictionary. The calculated sentiment values associated with females and males were then compared to identify patterns. The graphical representation of the distribution of overall sentiment (valence) scores associated with male and female genders are presented in Figure 2.

The sentiment values for males and females show considerable overlap with no significant difference in the distributions across negative and positive sentiments. However, we find that social media content targeting males indicates slightly more negative sentiments than positive sentiments. We therefore delved into gaining further insight into the emotions embedded within these sentiments. The following section expands on the emotion analysis results.

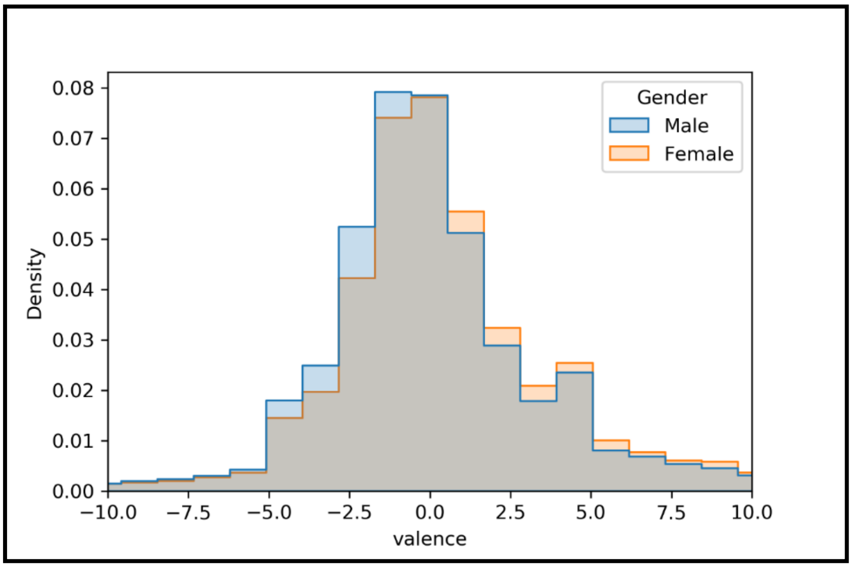


Figure 2. Sentiment value comparison

5.2.2 Emotion analysis

The eight primary emotions – joy, trust, anticipation, sadness, anger, fear, surprise and disgust – from Plutchik's wheel were analysed for the whole linguistic corpus. The 30000 comments/posts across both males and females showed the highest degrees of emotional expressions being in the form of disgust and sadness, with males receiving slightly more of the former and females more of the latter. The high-intensity values of the disgust category tended more towards loathing, while the sadness category tended towards grief. Our overall analysis reveals that females receive higher expressions of emotions generally in fat-based discussions on social media compared to males (see Table 3).

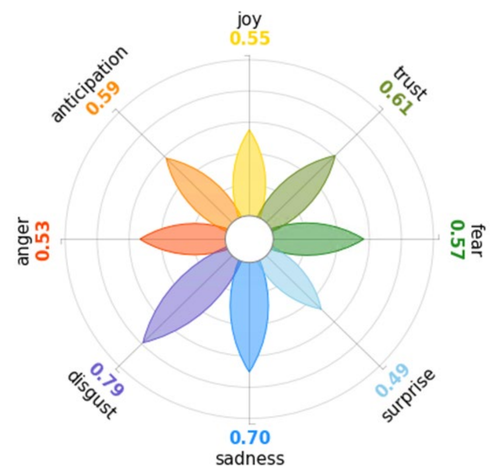
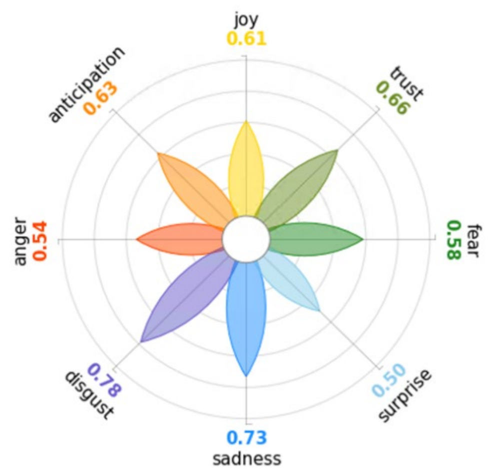
Plutchik Wheel of Emotions – Male	**Plutchik Wheel of Emotions - Female**
	

Table 3. Plutchik's wheel of emotions

The male and female emotions displayed in Plutchik's wheels represent huge amounts of social media big data spread across the eight segments. While data amounts used during analysis can be huge, researchers are advised to scale their data to a coarser level with a smaller subset of data and consider such visuals that make their presentation more readable to their audience (e.g., use bar charts for ranking and comparisons) (Kosara, 2016). Therefore, in Figure 3 we have highlighted the differences in the expressed emotions between males and females using percent differences. This figure depicts the percentage of overall differences between males and females that each emotion describes. Positive values describe emotions where the expression score was higher for females than males, and vice versa for negative values.

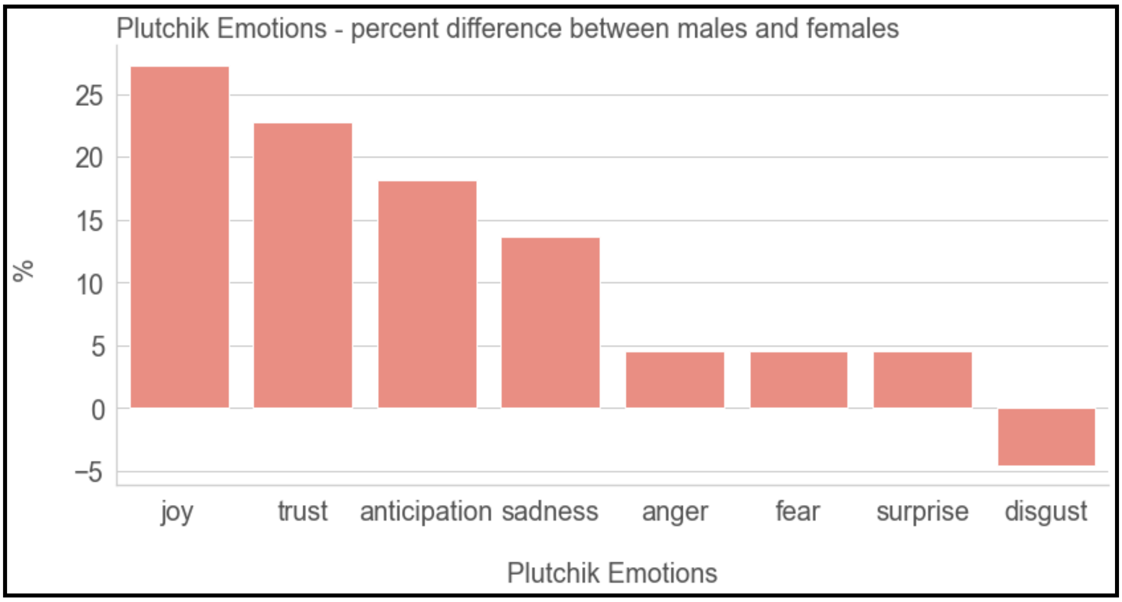


Figure 3. Percentage of all differences between males and females explained by each emotion

The figure indicates that some 25% of the total differences between emotions expressed between males and females can be explained by joy, where the females received a greater share. Indeed, the graph highlights that females received greater expressions of emotions across all categories, except for the disgust emotion which was more noticeable for males.

5.2.3 Sentiment labelling

To extract sentiment polarity, sentence-level sentiment analysis has been carried out next. Table 4 illustrates sentiment polarity labels (negative, neutral and positive) of three sentences which have been pre-processed (i.e., cleansed of stop words and other noisy characters). After obtaining the labels for each text-block, only the content which indicated negative sentiment labels was filtered and used as the input for pattern identification stage.

Cleaned Text	Valence	Sentiment
fat woman think theyre hot s**t please know die u	-1	negative
hit home skinny friend complain weight look like model	0	neutral
also im girl will will literally see beautiful plus size girl tv	1	positive

Table 4. Sample data for sentiment labelling

5.3 Pattern identification of fat stigma discussions

The underlying themes or patterns related to fat stigma have next been deduced with topic modelling approaches, word frequencies and discourse analysis to gain deeper insight on each topic. The following subsections elaborate on each of these methods.

5.3.1 Topic modelling

Topic models were generated by using Gensim LDA implementation for male and female datasets. The process outlined in Section 4.4 was used to determine the optimal number of topics for each dataset. The analysis yielded 5 topics for female dataset and 4 for the male dataset. Following the alpha and beta hyperparameter tuning for both models, Table 5 presents a before-and-after example of the degree to which the model coherence scores can be improved as a result of the model tuning process.

	Optimal Number of Topics	Coherence Score of Initial Model	Alpha	Beta	Coherence Score of Final Model
Male Dataset	4	0.36	0.61	0.01	0.66
Female Dataset	5	0.36	0.91	0.01	0.65

Table 5. Model parameters and coherence scores

After discovering the keywords generated in each topic, the authors reviewed these keywords to determine a suitable label for the topic. The set of keywords pertaining to a topic generated for each dataset along with topics' labels are displayed in Table 6. The derogatory and offensive words included in the generated topics are hidden by using the '*' character.

Fat stigma content targeting males	
Keywords of generated topic	**Topic label**
male, black, woman, good, fat, girl, man, want, feel, think	Opposite-sex
video, mom, time, d**k, never, eat, play, boy, call, watch	Lifestyle
f**k, say, gay, little, a*s, guy, thing, come, drink, still	Sexual
get, big, go, force, make, know, t*t, mind, also, way	Exercise
Fat stigma content targeting females	
Keywords of generated topic	**Topic label**
sex, woman, good, want, man, start, day, bad, new, sure	Relationships
get, go, make, say, take, see, guy, thing, never, leave	Lifestyle
f**k, big, girl, hot, p***y, c**k, a*s, mom, sexy, give	Sexual
even, know, look, fat, think, body, little, tell, also, still	Body
female, black, naked, time, love, people, old, mind, blonde, really	Appearance

Table 6. Generated topics and topic labels

5.3.2 Word Frequencies

The most prominent words of the corpus have been identified for both males and females in Figures 4 and 5. We noted more derogatory words were observed in the male dataset compared to the female dataset, although some of these derogatory words referred to obese/fat males in female terms (e.g., mom, woman, girl). The use of feminine words to insult obese/fat males could indicate body shaming attitudes towards females, or, in other words, this could also indicate that voicing body objectifying content targeting females is considered acceptable.

Male

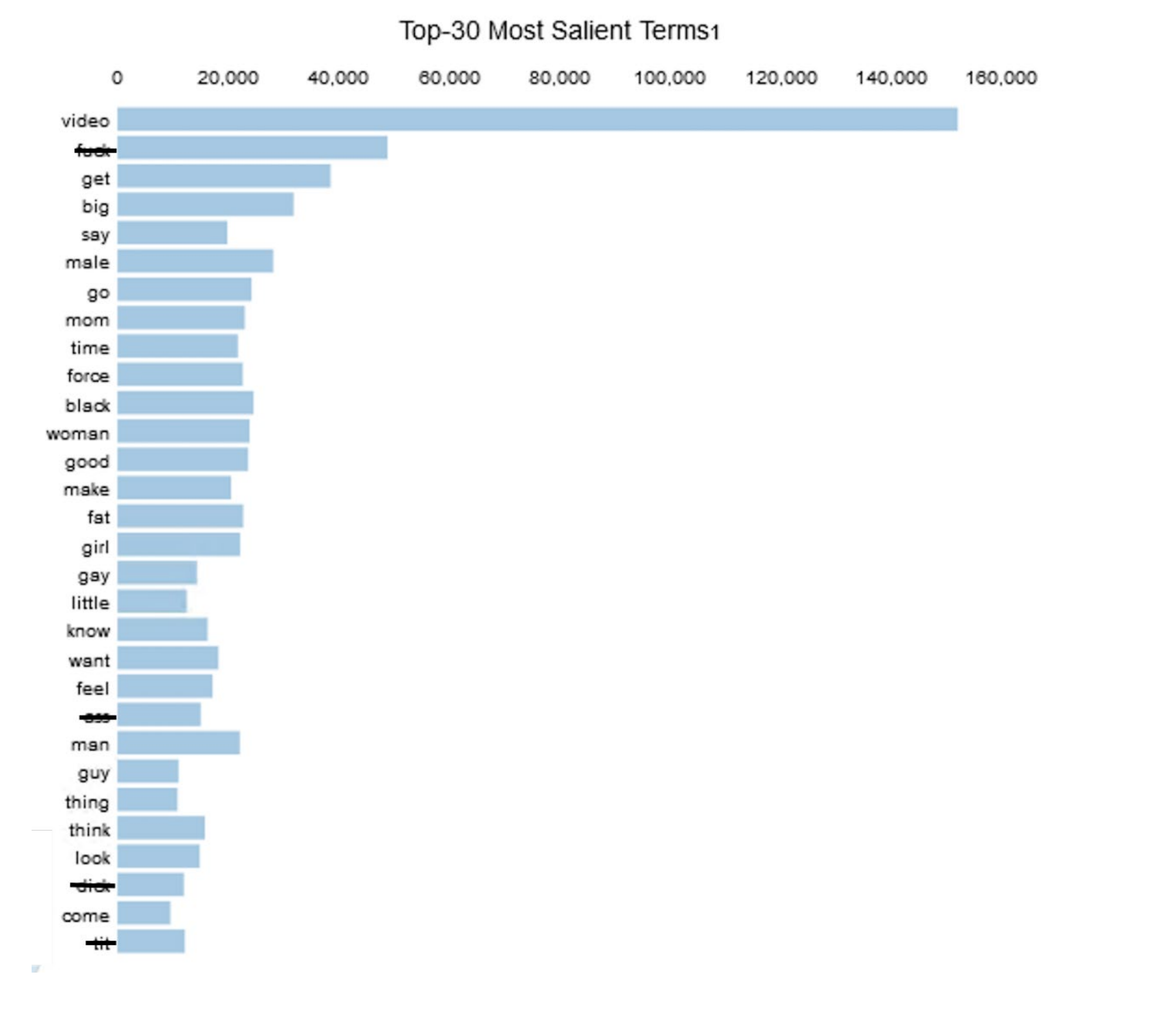


Figure 4. Word frequencies - Male

Female

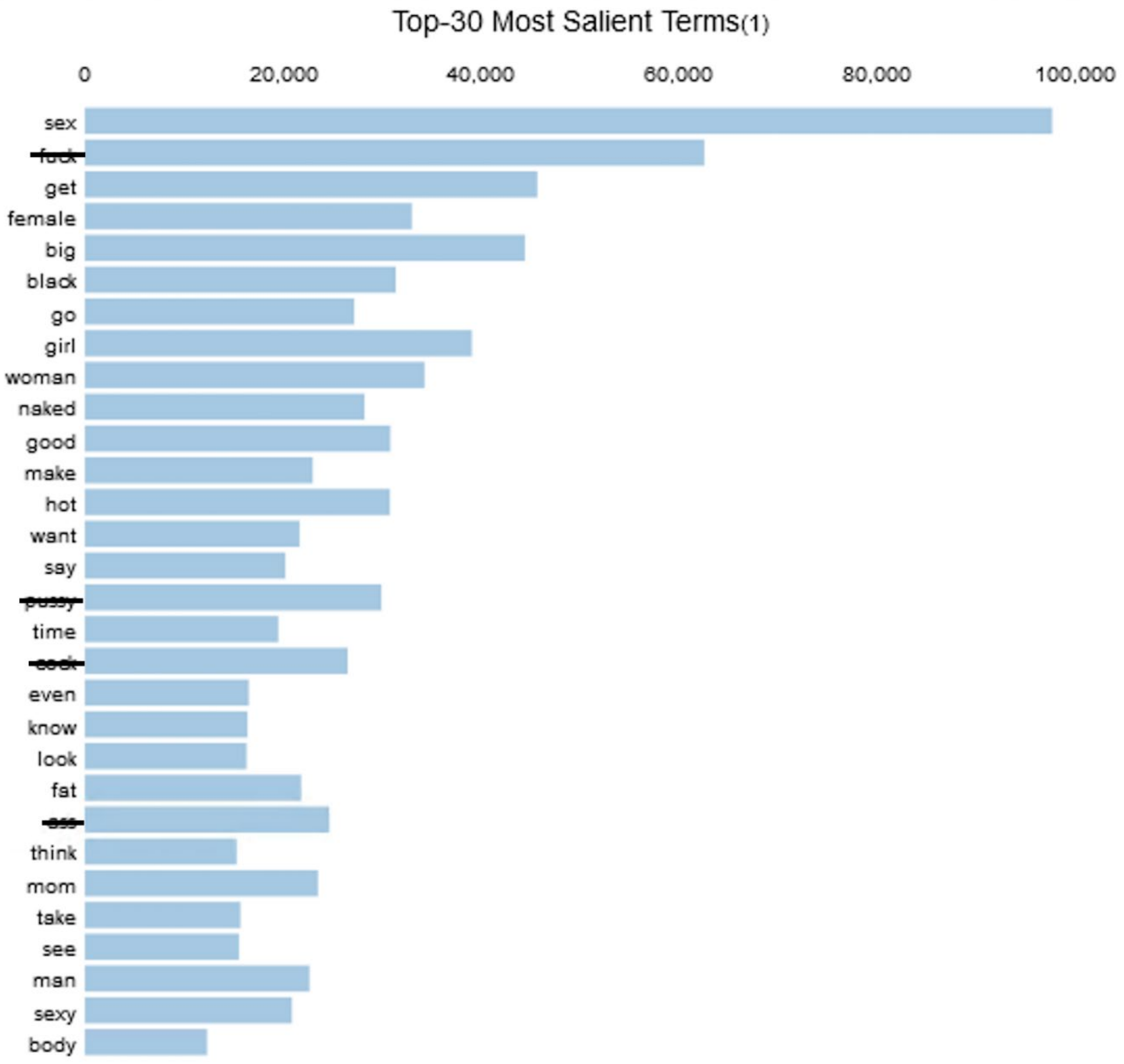


Figure 5. Word frequencies - Female

5.3.3 Discourse Analysis

Discourse analysis has been conducted by closely examining the most dominant comment/post for each of the identified topics. As a result, some more key features relating to fat stigma discussions were identified.

The fat stigma content targeting males indicates several features. For instance, some users shared scientific studies that they come across on social media; "*People with metabolically healthy obesity were at a substantially higher risk of diabetes, heart attack and stroke, heart failure, respiratory diseases, and all-cause mortality compared with people who were not obese and with a healthy metabolic profile*". Such content cannot be directly linked with fat stigma as the main purpose of sharing such information would be to educate readers. However, there is a possibility that a fat person may feel victimized based on such claims. Some other content revealed that the social media users consider fatness as 'unmasculine'; "*They are either smart, nerdy, short, fat, poor. Essentially unmasculine by society*". Such statements further indicate that some users consider their perspective as society's perspective. Individuals, who consider themselves obese, on reading such comments can adopt such statements and believe that this is how society sees them.

Obesity is found to be most commonly linked with diet and eating habits by social media users; "*ate fast food everyday (led to me becoming extremely obese for my age)*". Some even provided tips on losing weight; "*I've been the obese 13 year old that wanted to lose weight and I asked everyone what to do about it. My parents didn't know because they had bad eating habits themselves…. The only one who told me the secret recipe was a friend. "Well I just don't eat more than once a day, I only eat boiled veggies and I don't drink much water because it bloats you". And that was my diet for the next three years (13-16) and I lost all the weight reaching 50kgs/110lbs*". Such statements connect body weight with lifestyle choices and mostly try to motivate people to adopt 'healthy' lifestyles. Some users exhibited fear of rejection because of their fat bodies while at the same time indicating fat people are unattractive; "*To my horror I find myself being obese, bald, and unattractive. This shocked me, if I ever find my girlfriend, is this how I will look to her? How am I going to explain myself? Is she going to just dump me? And now, how am I going to reach her?*". These statements are directly linked with body objectification as the connection of body weight with attractiveness is established here. It further highlights that body objectification occurs in males as well as in females.

When targeting females on fat stigma, few social media users linked fatness with appearance; "*But I can confidently say she is not the sperm whale you make her sound like. From my general search, she is what some folks would call 'thick' which is not technically considered fat, as she does not have a belly sticking out*" and sometimes with health and well-being; "*your girlfriend is overweight, but she is not obese. Anytime you are overweight, there is some possible significant health risk, and it can be great that you care about her well-being*". These comments further highlight the connection of body weight with lifestyle choices. The dataset also revealed that users share scientific and statistical studies relating to obesity to educate the readers; "*5.8 percent of healthy-weight females with no sexual assault history attempted suicide. The percentage rose to 27.1 percent for healthy-weight girls with a history of sexual assault. Weight influenced the suicide rate among women: 8.2 percent of overweight girls with no sexual assault history attempt suicide*". However, compared to males, fat stigma content targeting females exposed a higher amount of sexual content. In most of these comments/posts, the words 'fat' and 'obese' were used to describe a person or a female body part, sexually. These contents were mainly posted either to share users' sexual preferences or to describe visual content.

6 Discussion

The study has proposed a methodological framework to examine hate speech patterns underpinning discussions in social media. Although sentiment analysis and topic modelling methods have been integrated into prior hate speech (or fat stigma) research, the application of such integration to identify patterns of hate speech is poorly discussed. Further, very few studies have incorporated natural language processing and machine learning techniques with discourse analysis for uncovering societal aspects that are associated with hate speech. Our methodological framework overcomes these limitations to show how sentiment analysis, topic modelling and discourse analysis can be applied to a given dataset. Section 6.1 presents deliberations on issues faced when integrating computer-mediated quantitative analytic tools with qualitative discourse analysis to provide a synthesis of multiple approaches for conducting data-driven analysis on large linguistic corpora. Next, in section 6.2, we discuss the framework's application in a practical context with a fat stigma investigation (using data extracted from Twitter and Reddit). Our analysis of obesity-related content with the selected methods has revealed the linguistic and thematic patterns associated with fat stigma content that is posted on social media conversations.

6.1 Application of the framework

The paper has proposed a methodological framework that shows the conceptual relevancy of different steps undertaken to detect hate speech patterns in social media. The framework details a four-step process that reveals specific activities within each step explains the conduct of each activity and provides insights on overcoming shortcomings with existing techniques, to report on data-driven findings. The first step of the proposed methodological framework is to identify the most relevant keywords from the published literature that align with the study's context. It may be noted, however, that the usage of jargon language has evolved; hence, the nature of online commenting specifically on social media too has changed (Eisenstein et al., 2014). Consequently, the linguistic awareness around words expressed in everyday conversations over social media platforms needs proper examination for currency and relevancy. Moreover, in collecting relevant social media data by conducting keyword searches, we are very likely to acquire the same comment/post multiple times. The Twitter API assisted in the elimination of retweets (or duplicate tweets) while scraping Twitter data. However, the duplicate comments/posts from Reddit need to be explicitly removed in the data pre-processing step. We advise that to avoid duplicates from Reddit, the data acquisition process considers data from a single subreddit since duplicates mainly occur due to posting the same content on multiple subreddits. Next, we have highlighted several methods for noise removal. By removing stop words, symbols, emojis and URLs, much of the noise in the text can be eliminated. However, in modern-day texting and social media postings, emojis or emoticons play a considerable role in conveying emotions. For example, "I had a rough day LLL" has more emotions than "I had a rough day". Emojis or emoticons can enhance digital communication by adding non-verbal cues (Alshenqeeti, 2016). Another prominent characteristic of digital text communication is amplifications or abbreviated words, commonly known as SMS language, textspeak, or texting language (Stieglitz & Dang-Xuan, 2013). For example, the canonical form of the words 'gooood' and 'gud' is 'good'. Normalizing social media text is important to achieve deeper text analysis, specifically on social media data. Satapathy et al. (2017) observed that the accuracy of sentiment classification has increased by ~4% when using such normalization. Therefore, it is clearly important to enhance available text mining techniques to cope with emojis and textspeak for more accurate text analysis.

After pre-processing the collected data, sentiment analysis and emotion analysis were undertaken. Patterns of sentiments and emotions that were scattered throughout the dataset were revealed, which provided knowledge of the most prominent emotions and sentiments related to a specific subject context. Further, sentiment analysis also facilitated in the filtering of the negative content, which led the analysis onward, since our purpose is to understand hate speech patterns that are exposed from negative sentiments. Next, the framework guided in in revealing patterns including hidden topics, frequent words and discourse patterns related to the identified negative content of the selected context. These patterns were inferred with the use of sentiment analysis and topic modelling on the corpus of texts and then interpreted qualitatively using discourse analysis (as described in section 4). In this manner, additional analyses can be incorporated to help in disclosing dominant features of hate speech expressions on social media. Our framework highlights the use of discourse analysis as an extension of computer-mediated tools and techniques to gain deeper insight into the hate content. Discourse analysis can be combined with topic modelling by connecting filtered comments/posts with the strength/size of the topics to thematise topic categories (Shirazi, 2013; Törnberg & Törnberg, 2016). The knowledge categories thus discovered can provide more societal background information on the specified subject to enable better interpretation (e.g., hate content relating to an incident that has occurred). When seeking hate speech patterns, we need to focus on who the targets of offensive expressions are, or which content or words cause the most impact, or are intense discussions occurring because of some event. The manual exploration in carrying out discourse analysis cannot be overlooked; hence having relevant data extracts are important. Therefore, technological solutions, such as sentiment analysis can aid in filtering out relevant posts/comments for attaining more accurate discourse analysis that can be used for larger datasets. The dataset acquired from keyword search needs to be further contextualized in the context of the study objective. Hence, to observe hate speech patterns, we must first remove the content that exhibits positive sentiments. The positive sentiments will otherwise act as noise in the detection of hate speech patterns. The content that indicates negative sentiments is carried forward for understanding the hate content. Finally, discourse analysis helps to provide a richer understanding of the subjective negative content, rather than labelling all content as offensive content.

Our framework provides the theoretical grounding for combining sentiment analysis with topic modelling, and subsequently conducting discourse analysis for identifying discussion themes/topics (e.g., fat hate speech) in social media discussions. The following section discusses the result of the data-driven analysis methods (outlined in our proposed framework) for the study of fat-stigma patterns.

6.2 Emergent fat-stigma themes

We have demonstrated the application of the methodological framework to investigate gender-based aspects related to fat-stigma discussions conducted over social media. Datasets, extracted from Twitter and Reddit, representing male and female genders across fat-related discussions were examined.

The study has disclosed noteworthy themes regarding the positioning of males and females from a fat stigma context. Sentiment analysis tools showed a mix of positive and negative sentiments for females and males. Emotion analysis has revealed the higher intensity of emotions being projected towards females than towards males. We found higher expressions relating to disgust (loathing) targeting males and of sadness (grief) targeting women. The emotion analysis results tuned the sentiment analysis outputs and highlighted that although males received a slightly higher number of positive sentiments, females tend to get more varied emotions compared to males. Such findings further justify the decision to use multiple text analysis methods to analyse hate speech.

Fat-based themes targeting females consisted of diet plans, exercise routines and health-related information, with most of these comments/posts having some degree of sexual content. Fat-related words were often used to describe female body parts in highly sexed manners. Further, these words were paired with other derogatory words that spread more negativity toward females. Moreover, males were often belittled by words that aligned with the female gender (e.g., lady, girl). This finding aligns with the objectification theory established by Fredrickson and Roberts (1997) which takes account of how female bodies are sexually evaluated based on "their bodies or body parts" (p. 176). Such body objectification can cause women to adopt an outsider view, where they equate their self-worth with their physical appearance and start self-objectifying and self-monitoring their body shapes. Objectification theory has been used as a theoretical framework for studying the experiential consequences of being a female in a culture which sexually objectifies the female body. (Tylka & Hill, 2004). The topics that emerged from this study provide rich insights into how weight prejudices are reflected in fat-based discussions. We found common topics from both male and female datasets. These include general categories including body, appearance, time, exercise, and sex. A closer examination of these categories further indicates the extent of sexual objectification. The sexual topics targeting males comprised words that have a strong association with the female gender (e.g., 'sex,video,girl,…', 'sl*t,big,wife,…'). While these findings were observed in a male context, the use of objectified female body parts (i.e., 'fat a*s', 'fat b**t') itself indicates further stigmatization of women. These fat stigma patterns illustrate the extent of female body objectification taking place and further demonstrate the application of our methodological framework for advancing the knowledge base and its alignment with current social theories.

7 Conclusions

This paper has presented a methodological framework, which has subsequently been applied to real-world conversations held over social media platforms. Specifically, we have studied hate speech patterns related to fat stigma in a gendered context to reveal behaviours. The proposed step-by-step framework provides researchers with methodological grounding on conducting natural language processing with social media data extracts. Our framework determines that the context of the problem domain needs to be articulated explicitly so that data-driven insights can be revealed. A key part of using automated analytics is that relevant data is collected and cleansed properly, following which the analysis is context-based to create real-world insights. Creating insights is not straightforward, hence we propose guidelines on the use of mixed approaches that include different data-driven as well as human-cantered qualitative analysis methods. Current information systems research faces methodological challenges in their use of multiple computational methods for studying large data corpus "to discover different patterns in their data, thereby highlighting different aspects of the phenomenon" which can then be further corroborated with traditional analytical methods (Miranda et al., 2022, p. xi). Our framework has established a roadmap for conducting mixed-method analyses for a better comprehension and understanding of conversational patterns (specifically in the context of negatively expressed content).

The framework takes hate speech detection a step forward by combining existing NLP techniques to investigate discussion patterns of fat stigma (or hate speech) in a social media context. It bridges the gap between technology and social sciences to enable a deeper understanding of the severity of the socio-psychological impact that fat stigma can cause on fat people. Though this study targets fat stigma, the suggested framework can be applied to various categories of hate speech to identify discussion patterns. Outputs of such pattern identification can assist in developing machine learning techniques for mitigating hate speech content on social media platforms or other public forums.

One of the limitations of this research is that the proposed tools relating to each method of the framework are solely based on literature and researchers' experience. Future studies could compare multiple tools relating to each method and obtain the tools that indicate the most effective results. The proposed framework is focused on text-based hate speech analysis, while hate speech can be observed in many forms, including visual and audio. In future, the research could expand this framework to incorporate methods of visual and audio analysis to overcome this limitation. Also, the application of the framework has been tested on Twitter and Reddit. However, various types of text-based online media such as forums and weblogs are currently available. Therefore, in future, studies could be conducted to identify patterns of hate speech (fat stigma) on these online media platforms.

This paper has demonstrated the development of analytical social patterns from real-world conversations by using computational and replicable design methods; therefore, it has implications for conducting more forms of interdisciplinary research. We encourage future research studies to use our proposed methodological framework for investigating societal issues, such as examining online comments posted on social media sites to detect if people are being negatively targeted based on their body image, ethnic background, gender, religion or disabilities amongst similar others.

References

al-Utbi, M. I. K. (2019). A Critical Discourse Analysis of Hate Speech. *Journal of the College of Languages (JCL) Mağallaẗ kulliyyaẗ al-luġāt* (39), 19-40. https://doi.org/10.36586/jcl.2.2019.0.39.0019

Albadi, N., Kurdi, M., & Mishra, S. (2018). *Are they our brothers? analysis and detection of religious hate speech in the arabic twittersphere*. Paper presented at the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).Barcelona, Spain. 28 – 31 August 2018. https://dl.acm.org/doi/10.5555/3382225.3382239

Almenara, C. A., & Ježek, S. (2015). The source and impact of appearance teasing: an examination by sex and weight status among early adolescents from the Czech Republic. *Journal of School Health, 85*(3), 163-170. doi: https://doi.org/10.1111/josh.12236

Alshalan, R., Al-Khalifa, H., Alsaeed, D., Al-Baity, H., & Alshalan, S. (2020). Detection of Hate Speech in COVID-19–Related Tweets in the Arab Region: Deep Learning and Topic Modeling Approach. *Journal of Medical Internet Research, 22*(12), e22609. https://doi.org/10.2196/22609

Alshenqeeti, H. (2016). Are emojis creating a new or old visual language for new generations? A socio-semiotic study. *Advances in Language and Literary Studies, 7*(6). 56-69. https://doi.org/10.31219/osf.io/4hdgs

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993-1022.

Boban, I., Doko, A., & Gotovac, S. (2020). Sentence retrieval using stemming and lemmatization with different length of the queries. *Advances in Science, Technology and Engineering Systems, 5*(3), 349-354. https://doi.org/10.25046/aj050345

Boettcher, N. (2021). Studies of Depression and Anxiety Using Reddit as a Data Source: Scoping Review. *Journal of Medical Internet Research Mental Health, 8*(11), e29487. https://doi.org/10.2196/29487

Bograd, S., Chen, B., & Kavuluru, R. (2022). Tracking sentiments toward fat acceptance over a decade on Twitter. *Health Informatics Journal, 28*(1), 1-16. https://doi.org/10.1177/14604582211065702

Brewis, A., SturtzSreetharan, C., & Wutich, A. (2018). Obesity stigma as a globalizing health challenge. *Globalization and Health, 14*(1), 1-6. https://doi.org/10.1186/s12992-018-0337-x

Brooker, P., Barnett, J., Vines, J., Lawson, S., Feltwell, T., & Long, K. (2018). Doing stigma: Online commenting around weight-related news media. *New Media & Society, 20*(9), 3201-3222. https://doi.org/10.1177/1461444817744790

Calderón, C. A., de la Vega, G., & Herrero, D. B. (2020). Topic Modeling and Characterization of Hate Speech against Immigrants on Twitter around the Emergence of a Far-Right Party in Spain. *Social Sciences, 9*(11), 188. https://doi.org/10.3390/socsci9110188

Cao, R., Lee, R. K.-W., & Hoang, T.-A. (2020). *DeepHate: Hate speech detection via multi-faceted text representations*. Paper presented at the 12th ACM Conference on Web Science. Southampton, United Kingdom, 6 – 10 July 2020. https://dl.acm.org/doi/10.1145/3394231.3397890

Chiril, P., Moriceau, V., Benamara, F., Mari, A., Origgi, G., & Coulomb-Gully, M. (2020). *He said "who's gonna take care of your children when you are at ACL?": Reported Sexist Acts are Not Sexist*. Paper presented at the Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 4055-4066. Online. Association for Computational Linguistics. https://aclanthology.org/2020.acl-main.373.pdf

Chou, W. Y. S., Prestin, A., & Kunath, S. (2014). Obesity in social media: a mixed methods analysis. *Translational Behavioral Medicine, 4*(3), 314-323. https://doi.org/10.1007/s13142-014-0256-1

Curiskis, S. A., Drake, B., Osborn, T. R., & Kennedy, P. J. (2020). An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. *Information Processing & Management, 57*(2), 102034. https://doi.org/10.1016/j.ipm.2019.04.002

De Brún, A., McCarthy, M., McKenzie, K., & McGloin, A. (2014). Weight stigma and narrative resistance evident in online discussions of obesity. *Appetite, 72*, 73-81. https://doi.org/10.1016/j.appet.2013.09.022

Delgado, R., & Stefancic, J. (1991). Images of the outsider in American law and culture: Can free expression remedy systemic social ills. *Cornell Law Review, 77*, 1258. https://doi.org/10.4324/9780429037627-2

Delgado, R., & Stefancic, J. (2014). Hate speech in cyberspace. *Wake Forest Law Review, 49*, 319. https://scholarship.law.ua.edu/fac_articles/560

Durrheim, K., Okuyan, M., Twali, M. S., García-Sánchez, E., Pereira, A., Portice, J. S., Gur, T., Weiner-Blotmer, O. & Keil, T. F. (2018). How racism discourse can mobilize right-wing populism: The construction of identity and alliance in reactions to UKIP's Brexit "Breaking Point" campaign. *Journal of Community & Applied Social Psychology, 28*(6), 385-405. https://doi.org/10.1002/casp.2347

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2014). Diffusion of lexical change in social media. *PloS one, 9*(11), e113114. https://doi.org/10.1371/journal.pone.0113114

Erjavec, K., & Kovačič, M. P. (2012). "You Don't Understand, This is a New War!" Analysis of Hate Speech in News Web Sites' Comments. *Mass Communication and Society, 15*(6), 899-920. https://doi.org/10.1080/15205436.2011.619679

Fredrickson, B. L., & Roberts, T. A. (1997). Objectification theory: Toward understanding women's lived experiences and mental health risks. *Psychology of Women Quarterly, 21*(2), 173-206. https://doi.org/10.1111/j.1471-6402.1997.tb00108.x

Gallacher, J. D., Heerdink, M. W., & Hewstone, M. (2021). Online engagement between opposing political protest groups via social media is linked to physical violence of offline encounters. *Social Media+ Society, 7*(1), https://doi.org/10.1177/2056305120984445

Himmelstein, M. S., Puhl, R. M., & Quinn, D. M. (2017). Intersectionality: an understudied framework for addressing weight stigma. *American Journal of Preventive Medicine, 53*(4), 421-431. https://doi.org/10.1016/j.amepre.2017.04.003

Holmberg, C., Berg, C., Hillman, T., Lissner, L., & Chaplin, J. E. (2018). Self-presentation in digital media among adolescent patients with obesity: Striving for integrity, risk-reduction, and social recognition. *Digital Health, 4*. https://doi.org/10.1177/2055207618807603

Hussin, M., Frazier, S., & Thompson, J. K. (2011). Fat stigmatization on YouTube: A content analysis. *Body Image, 8*(1), 90-92. https://doi.org/10.1016/j.bodyim.2010.10.003

Irani, A., Hendry, Manongga, D. H. F., & Chen, R.-C. (2020). Mining Public Opinion on Radicalism in Social Media Via Sentiment Analysis. *International Journal of Innovative Computing, Information and Control, 16*(5). 1787-1800. https://doi.org/10.24507/ijicic.16.05.1787

Izydorczyk, B., Walenista, W., Kamionka, A., Lizińczyk, S., & Ptak, M. (2021). Connections Between Perceived Social Support and the Body Image in the Group of Women With Diastasis Recti Abdominis. *Frontiers in Psychology, 12*. 3182. https://doi.org/10.3389/fpsyg.2021.707775

Jeon, Y. A., Hale, B., Knackmuhs, E., & Mackert, M. (2018). Weight stigma goes viral on the Internet: Systematic assessment of YouTube comments attacking overweight men and women. *Interactive Journal of Medical Research, 7*(1), e6. https://doi.org/10.2196/ijmr.9182

Kemp, S. (2021). *Digital 2021: Global Overview Report*. Retrieved from https://datareportal.com/reports/digital-2021-global-overview-report. Accessed 30 August 2022.

Kent, E. E., Prestin, A., Gaysynsky, A., Galica, K., Rinker, R., Graff, K., & Chou, W.-Y. S. (2016). “Obesity is the new major cause of cancer”: connections between obesity and cancer on facebook and twitter. *Journal of Cancer Education, 31*(3), 453-459. https://doi.org/10.1007/s13187-015-0824-1

Kocoń, J., Figas, A., Gruza, M., Puchalska, D., Kajdanowicz, T., & Kazienko, P. (2021). Offensive, aggressive, and hate speech analysis: From data-centric to human-centered approach. *Information Processing & Management, 58*(5), 102643. https://doi.org/10.1016/j.ipm.2021.102643

Kosara, R. (2016). Presentation-oriented visualization techniques. *IEEE Computer Graphics and Applications, 36*(1), 80-85. https://doi.org/10.1109/mcg.2016.2

Larson, S. R. (2021). The Rhetoricity of Fat Stigma: Mental Disability, Pain, and Anorexia Nervosa. *Rhetoric Society Quarterly, 51*(5), 392-406. https://doi.org/10.1080/02773945.2021.1972131

Lazarus, J. V., Kakalou, C., Palayew, A., Karamanidou, C., Maramis, C., Natsiavas, P., Picchio, C.A., Villota-Rivas, M. Zelber-Sagi, S. & Carrieri, P. (2021). A Twitter discourse analysis of negative feelings and stigma related to NAFLD, NASH, and obesity. *Liver International, 41*(10), 2295-2307. https://doi.org/10.1111/liv.14969

Lipizzi, C., Iandoli, L., & Marquez, J. E. R. (2015). Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using Twitter streams. *International Journal of Information Management, 35*(4), 490-503. https://doi.org/10.1016/j.ijinfomgt.2015.04.001

Liu, B. (2012). *Sentiment analysis and opinion mining* (Vol. 5). Morgan & Claypool Publishing, Springer Chan. https://doi.org/10.1007/978-3-031-02145-9

Lydecker, J. A., Cotter, E. W., Palmberg, A. A., Simpson, C., Kwitowski, M., White, K., & Mazzeo, S. E. (2016). Does this Tweet make me look fat? A content analysis of weight stigma on Twitter. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 21*(2), 229-235. https://doi.org/10.1007/s40519-016-0272-x

Maftei, A., & Merlici, I.-A. (2022). Am I thin enough? Social media use and the ideal body stereotype: The mediating role of perceived socio-cultural pressure and the moderating role of cognitive fusion. *Current Psychology*, 1-14. https://doi.org/10.1007/s12144-022-02938-x

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *An Introduction to Information Retrieval*. Cambridge, England: Cambridge University Press.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014). *The Stanford CoreNLP natural language processing toolkit*. Paper presented at the Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations. 55–60, Baltimore, Maryland. Association for Computational Linguistics. https://aclanthology.org/P14-5010

Martins, R., Gomes, M., Almeida, J. J., Novais, P., & Henriques, P. (2018). *Hate speech classification in social media using emotional analysis*. Paper presented at the 2018 7th Brazilian Conference on Intelligent Systems (BRACIS). 22-25 October 2018, Sao Paulo, Brazil. 61-66. https://doi.org/10.1109/BRACIS.2018.00019

Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & New Media, 22*(2), 205-224. https://doi.org/10.1177/1527476420982230

Mathew, B., Illendula, A., Saha, P., Sarkar, S., Goyal, P., & Mukherjee, A. (2020). Hate begets hate: A temporal study of hate speech. *Proceedings of the ACM on Human-Computer Interaction, 4*(CSCW2), 92. 1-24. https://doi.org/10.1145/3415163

McComb, S. E., & Mills, J. S. (2022). The effect of physical appearance perfectionism and social comparison to thin-, slim-thick-, and fit-ideal Instagram imagery on young women's body image. *Body Image, 40*, 165-175. https://doi.org/10.1016/j.bodyim.2021.12.003

Microsoft Corporation. (2016). Microsoft Excel (Version 2209) [Computer Software]. Retrieved from https://www.microsoft.com/en-nz/microsoft-365/excel. Accessed 01 January 2022.

Miranda, S., Berente, N., Seidel, S., Safadi, H., & Burton-Jones, A. (2022). Editor's Comments: Computationally Intensive Theory Construction: A Primer for Authors and Reviewers. *Management Information Systems Quarterly*, *46*(2), iii-xviii.

Mohammad, S. M., & Turney, P. D. (2013). NRC emotion lexicon. *National Research Council Canada*, Technical report (NPARC number: 21270984), 243p. https://doi.org/10.4224/21270984. Accessed 31 May 2022.

Mondal, M., Silva, L. A., & Benevenuto, F. (2017). *A measurement study of hate speech in social media*. Paper presented at the Proceedings of the 28th ACM conference on hypertext and social media. 4-7 July 1017, Prague, Czech Republic. 85-94. https://doi.org/10.1145/3078714.3078723

Mossie, Z., & Wang, J.-H. (2020). Vulnerable community identification using hate speech detection on social media. *Information Processing & Management, 57*(3), 102087. https://doi.org/10.1016/j.ipm.2019.102087

Murakami, J. M., & Latner, J. D. (2015). Weight acceptance versus body dissatisfaction: Effects on stigma, perceived self-esteem, and perceived psychopathology. *Eating Behaviors, 19*, 163-167. https://doi.org/10.1016/j.eatbeh.2015.09.010

Ophir, Y., Walter, D., & Marchant, E. R. (2020). A collaborative way of knowing: Bridging computational communication research and grounded theory ethnography. *Journal of Communication, 70*(3), 447-472. https://doi.org/10.1093/joc/jqaa013

Pronoza, E., Panicheva, P., Koltsova, O., & Rosso, P. (2021). Detecting ethnicity-targeted hate speech in Russian social media texts. *Information Processing & Management, 58*(6), 102674. https://doi.org/10.1016/j.ipm.2021.102674

Puhl, R. M., Andreyeva, T., & Brownell, K. D. (2008). Perceptions of weight discrimination: prevalence and comparison to race and gender discrimination in America. *International Journal of Obesity, 32*(6), 992. https://doi.org/10.1038/ijo.2008.22

Rawal, K., Patel, T. P., Purohit, K. M., Israni, K., Kataria, V., Bhatt, H., & Gupta, S. (2020). Influence of obese phenotype on metabolic profile, inflammatory mediators and stemness of hADSC in adipose tissue. *Clinical Nutrition, 39*(12), 3829-3835. https://doi.org/10.1016/j.clnu.2020.02.032

Reddit. (2020). Reddit API Documentation. Retrieved from https://www.reddit.com/dev/api/. Accessed 01 Jan 2021.

Řehůřek, R., & Sojka, P. (2011). Gensim—statistical semantics in python. Poster presented at EuroScipy 2011, Paris, 25.–28. 8. 2011. https://initiative.eudml.org/news/presentation-euroscipy-2011-paris. Accessed 01 Jan 2022.

Röder, M., Both, A., & Hinneburg, A. (2015). *Exploring the space of topic coherence measures*. Paper presented at the Proceedings of the eighth ACM international conference on Web search and data mining, February 2 -6, 2015, Shanghai, China. 399-408. https://doi.org/10.1145/2684822.2685324

Rodríguez, A., Argueta, C., & Chen, Y.-L. (2019). *Automatic detection of hate speech on facebook using sentiment and emotion analysis*. Paper presented at the 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). 11-13 February 2019, Okinawa, Japan. https://ieeexplore.ieee.org/document/8669073

Salminen, J., Almerekhi, H., Milenković, M., Jung, S.-g., An, J., Kwak, H., & Jansen, B. J. (2018). *Anatomy of online hate: developing a taxonomy and machine learning models for identifying and classifying hate in online news media*. Paper presented at the Twelfth International AAAI Conference on Web and Social Media. 330-339. 25 -28 June 2018. Palo Alto, United States. https://pennstate.pure.elsevier.com/en/publications/anatomy-of-online-hate-developing-a-taxonomy-and-machine-learning

Satapathy, R., Guerreiro, C., Chaturvedi, I., & Cambria, E. (2017). *Phonetic-based microtext normalization for twitter sentiment analysis*. Paper presented at the 2017 IEEE international conference on data mining workshops (ICDMW). 18-21 November 2017. New Orleans, LA, USA https://doi.org/10.1109/ICDMW.2017.59

Semeraro, A., Vilella, S., & Ruffo, G. (2021). PyPlutchik: Visualising and comparing emotion-annotated corpora. *PloS one, 16*(9), e0256503. https://doi.org/10.1371/journal.pone.0256503

Shibly, F., Sharma, U., & Naleer, H. (2021). *Classifying and measuring hate speech in Twitter using topic classifier of sentiment analysis*. Paper presented at the International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, vol 1165. Springer, Singapore. 671-678. https://doi.org/10.1007/978-981-15-5113-0_54

Shirazi, F. (2013). Social media and the social movements in the Middle East and North Africa: A critical discourse analysis. *Information Technology & People. 26*(1). 28-49. https://doi.org/10.1108/09593841311307123

So, J., Prestin, A., Lee, L., Wang, Y., Yen, J., & Chou, W.-Y. S. (2016). What do people like to "share" about obesity? A content analysis of frequent retweets about obesity on Twitter. *Health Communication, 31*(2), 193-206. https://doi.org/10.1080/10410236.2014.940675

Ștefăniță, O., & Buf, D.-M. (2021). Hate speech in social media and its effects on the LGBT community: A review of the current research. *Romanian Journal of Communication and Public Relations, 23*(1), 47-55. https://doi.org/10.21018/rjcpr.2021.1.322

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems, 29*(4), 217-248. https://doi.org/10.2753/MIS0742-1222290408

Szepietowska, B., Polonsky, B., Sherazi, S., Biton, Y., Kutyifa, V., McNitt, S., . . . Zareba, W. (2016). Effect of obesity on the effectiveness of cardiac resynchronization to reduce the risk of first and recurrent ventricular tachyarrhythmia events. *Cardiovascular Diabetology, 15*(1), 1-8. https://doi.org/10.1186/s12933-016-0401-x

Tang-Péronard, J., & Heitmann, B. (2008). Stigmatization of obese children and adolescents, the importance of gender. *Obesity Reviews, 9*(6), 522-534. https://doi.org/10.1111/j.1467-789x.2008.00509.x

Thompson, L., Rickett, B., & Day, K. (2018). Feminist relational discourse analysis: Putting the personal in the political in feminist research. *Qualitative Research in Psychology, 15*(1), 93-115. https://doi.org/10.1080/14780887.2017.1393586

Tomiyama, A. J. (2014). Weight stigma is stressful. A review of evidence for the Cyclic Obesity/Weight-Based Stigma model. *Appetite, 82*, 8-15. https://doi.org/10.1016/j.appet.2014.06.108

Törnberg, A., & Törnberg, P. (2016). Muslims in social media discourse: Combining topic modeling and critical discourse analysis. *Discourse, Context & Media, 13*, 132-142. https://doi.org/10.1016/j.dcm.2016.04.003

Tucey, C. B. (2010). *Online vs. face-to-face deliberation on the global warming and stem cell issues.* Paper presented at the Western Political Science Association 2010 Annual Meeting Paper. https://ssrn.com/abstract=1580573

Twitter. (2019). Twitter API. Retrieved from https://developer.twitter.com/. Accessed 01 Jan 2021.

Tylka, T. L., & Hill, M. S. (2004). Objectification theory as it relates to disordered eating among college women. *Sex Roles, 51*(11-12), 719-730. https://doi.org/10.1007/s11199-004-0721-2

Veldhuis, J., Konijn, E. A., & Seidell, J. C. (2014). Negotiated media effects. Peer feedback modifies effects of media's thin-body ideal on adolescent girls. *Appetite, 73*, 172-182. https://doi.org/10.1016/j.appet.2013.10.023

Vigna, F. D., Cimino, A., Dell'Orletta, F., Petrocchi, M., & Tesconi, M. (2017). *Hate me, hate me not: Hate speech detection on facebook*. Paper presented at the Proceedings of the First Italian Conference on Cybersecurity (ITASEC17). Venice, Italy. 86- 95. https://ceur-ws.org/Vol-1816/paper-09.pdf

Wallach, H. M. (2006). *Topic modeling: beyond bag-of-words*. Paper presented at the Proceedings of the 23rd international conference on Machine learning. Pittsburgh, Pennsylvania. 25-29 July 2006. 977-984. https://doi.org/10.1145/1143844.1143967

Wanniarachchi, V. U., Mathrani, A., Susnjak, T., & Scogings, C. (2019). *Gendered objectification of weight stigma in social media: a mixed method analysis*. Paper presented at the Australasian Conference on Information Systems, Perth, Western Australia. 9-11 December 2019 https://acis2019.io/pdfs/ACIS2019_PaperFIN_077.pdf

Wanniarachchi, V. U., Mathrani, A., Susnjak, T., & Scogings, C. (2022). Methodological Aspects in Study of Fat Stigma in Social Media Contexts: A Systematic Literature Review. *Applied Sciences, 12*(10), 5045. https://doi.org/10.3390/app12105045

WHO. (2022). Data and statistics. Retrieved from https://www.euro.who.int/en/health-topics/noncommunicable-diseases/obesity/data-and-statistics. Accessed 01 Jan 2022.

Yeruva, V. K., Junaid, S., & Lee, Y. (2019). Contextual Word Embeddings and Topic Modeling in Healthy Dieting and Obesity. *Journal of Healthcare Informatics Research, 3*(2), 159-183. https://doi.org/10.1007/s41666-019-00052-5

Zhong, B. (2020). Social consequences of internet civilization. *Computers in Human Behavior, 107*, 106308. https://doi.org/10.1016/j.chb.2020.106308

doi: https://doi.org/10.3127/ajis.v27i0.3929