The Impact of COVID-19 Pandemic on LGBTQ Online Communities

Yunhao Yuan  
Aalto University  
Finland  
yunhao.yuan@aalto.fi

Gaurav Verma  
Georgia Institute of Technology  
United States  
gverma@gatech.edu

Barbara Keller  
Aalto University  
Finland  
barbara.keller@aalto.fi

Talayeh Aledavood  
Aalto University  
Finland  
talayeh.aledavood@aalto.fi

Abstract

The COVID-19 pandemic has disproportionately impacted the lives of minorities, such as members of the LGBTQ community (lesbian, gay, bisexual, transgender, and queer) due to pre-existing social disadvantages and health disparities. Although extensive research has been carried out on the impact of the COVID-19 pandemic on different aspects of the general population’s lives, few studies are focused on the LGBTQ population. In this paper, we identify a group of Twitter users who self-disclose to belong to the LGBTQ community. We develop and evaluate two sets of machine learning classifiers using a pre-pandemic and a during-pandemic dataset to identify Twitter posts exhibiting minority stress, which is a unique pressure faced by the members of the LGBTQ population due to their sexual and gender identities. For this task, we collect a set of 20,593,823 posts by 7,241 self-disclosed LGBTQ users and annotate a randomly selected subset of 2800 posts. We demonstrate that our best pre-pandemic and during-pandemic models show strong and stable performance for detecting posts which contain minority stress. We investigate the linguistic differences in minority stress posts across pre- and during-pandemic periods. We find that anger words are strongly associated with minority stress during the COVID-19 pandemic. We explore the impact of the pandemic on the emotional states of the LGBTQ population by conducting controlled comparisons with the general population. We adopt propensity score-based matching to perform a causal analysis. The results show that the LGBTQ population have a greater increase in the usage of cognitive words and worsened observable attribute in the usage of positive emotion words than the group of the general population with similar pre-pandemic behavioral attributes. Our findings have implications for the public health domain and policy-makers to provide more adequate support, especially with respect to mental health, to the LGBTQ population during future crises.

Introduction

The onset of the COVID-19 pandemic has profoundly impacted health, economy, and societies as a whole. Across the world, a wide range of substantive strategies, such as self-isolation, social distancing, and school closures, have been implemented to mitigate the spread of the coronavirus. These policies, which are effective for slowing down the spread of the virus, have come with extra costs to the mental well-being of individuals and communities.

Isolation from social relationships and disengagement from social activities can lead to adverse mental health outcomes such as depression, anxiety, and loneliness (Pietra-bissa and Simpson 2020). Recent work that studies these effects during the COVID-19 pandemic shows a considerable increase in psychological expressions, including expressions related to mental health issues and support, among the general population (Saha et al. 2020, Gallagher et al. 2020). The authors point out major concerns about personal challenges, health issues, and uncertainties surrounding the pandemic. The adverse effects of these introduced policies are shown to have a disproportional impact on different groups and communities within the society. For example, recent work by Pierce et al. (2020) assesses the mental distress level in the UK population before and during the lockdowns. Their research indicates deteriorated mental health during the pandemic, especially among women, youth, and kids.

Due to pre-existing social and structural inequalities, the COVID-19 pandemic is expected to have a greater impact on minority groups, including lesbian, gay, bisexual, transgender, and queer (LGBTQ) individuals. Even in the best of times, LGBTQ community members have experienced worse mental health outcomes resulting from persistent discrimination and stigmatization (Wang et al. 2020). This community struggles with health disparities, such as higher occurrences of long-term chronic illnesses and unique healthcare needs, compared to the straight and cisgender population (Zeeman et al. 2019). During the COVID-19 pandemic, Kneale and Becares (2020) report an increased level of stress and depressive symptoms among the LGBTQ community, particularly among transgender people. It is suggested that the closing of educational institutions and mandatory isolation during the pandemic may expose young members of the LGBTQ community to verbal abuse and domestic violence (Salerno, Williams, and Gattamorta 2020). However, research to date mainly relies on diagnostic reports and psychological questionnaires to analyze the mental health of the LGBTQ population. Due to the fear of cultural discrimination and stigmatization, a large number of LGBTQ individuals are hard to reach, limiting generaliz-
ability. To fill this gap, our study aims to investigate the mental health of the LGBTQ population during the pandemic by leveraging a large sample of social media data.

Social media are known to provide spaces where individuals and communities with stigmatized identities can express themselves, talk about their experiences and seek support, with less fear of offline harm (Andalibi et al. 2016). Data collected from different social media platforms have already been used to investigate stigmatized experiences in various domains such as drug misuse (Garg et al. 2021), sexual abuse (Andalibi et al. 2016), and mental health (Coppsmith et al. 2018). Similarly, social media offers an easily accessible and semi-anonymous avenue for LGBTQ individuals to connect with others, seek support and share their concerns (Cannon et al. 2017). Consequently, social media data provide an unparalleled opportunity to study the concerns of the LGBTQ community and their evolution over time. Harnessing this opportunity is especially important in the context of the COVID-19 pandemic, when mental health issues are known to have grown significantly.

In this work, we investigate the impact of COVID-19 on the LGBTQ community in the United States via Twitter data. We study minority stress which is a type of stress faced by the LGBTQ community members due to their sexual and gender identities. We address the following three research questions to better understand shifts in the minority stress patterns before and after the breakout of the COVID-19 pandemic and its disproportionate impacts on the LGBTQ population as compared to the general population:

RQ1: How can we measure minority stress on social media using computational methods?

RQ2: Are there linguistic differences in Twitter posts (tweets) containing minority stress in the time prior to the pandemic compared to minority stress posts during the pandemic?

RQ3: Are there linguistic differences in tweets between the LGBTQ and the general population during the pandemic?

To answer these research questions, we collect 20,593,823 tweets from 7,241 US-based self-disclosing LGBTQ Twitter users from June 2018 to June 2021. We manually annotate 2,800 tweets and divide them into two datasets. One set contains tweets posted before the COVID-19 pandemic, and the other set includes tweets posted during the pandemic. We train multiple machine learning classifiers separately on pre- and during-pandemic datasets, and evaluate the best performing classifiers on different feature sets. Our best pre-pandemic and during-pandemic models show stable and robust performance with an area under the receiver operating characteristic (AUC) curve of 0.838 and 0.845, respectively. We investigate the top statistically significant features of the pre-pandemic and during-pandemic models to compare the emotions behind the minority stress posts on Twitter. We find that anger has become strongly associated with minority stress during the COVID-19 pandemic. Lastly, we apply a stratified propensity score analysis to compare the COVID-19 pandemic effects on the LGBTQ population and the general population on social media. We observe that the COVID-19 pandemic has resulted in a higher usage of cognitive words and a reduced usage of positive emotion words for the LGBTQ population.

Privacy and Ethics

Despite the public accessibility of Twitter data, we adopt several approaches to protect the privacy of the data owners. We remove any information related to personal identity and paraphrase all quotations in this paper to avoid traceability. One of the authors identifies as a member of the LGBTQ community. This helps us better understand the privacy needs and concerns of this community.

Background and Related Work

Social Media Use During the COVID-19 Pandemic

Due to the need for social distancing during the COVID-19 pandemic, people increasingly use social media not only to consume and share information, but also to engage in seeking help, advice, and providing supportive information (Saad, Mashud, and Ida 2020). Recent work by Tsao et al. (2021) highlights various facets of public health research of social media during the pandemic, such as psychological reactions, government responses, and public attitudes.

Xiang et al. (2021) collect data from Twitter and Decahose to analyze the public discourse and sentiments relative to older people and the extent of ageism among social media users. Zhou et al. (2021) build a depression classifier by extracting multi-modal features from Twitter data, such as emotion, topic, and domain-specific features, and analyze the local dynamics during the COVID-19 in Australia. Verma et al. (2022) investigate the 80 million Twitter posts and demonstrate that users sharing COVID-19 related misinformation will experience exacerbated anxiety.

LGBTQ Community on Social Media

Social media plays a prominent role in the exchange of information. It also provides opportunities and platforms for minorities and communities such as LGBTQ individuals to engage in online activities, such as seeking out social support, exchanging information, and maintaining social relationships (Steinke et al. 2017). Prior research suggests that LGBTQ youths use social media as a primary method of socializing and spend more time online than their counterpart cisgender individuals (Steinke et al. 2017).

Given the significant growth in social media usage, data from LGBTQ online communities has been used to analyze social media’s effects on the LGBTQ population. One study (Craig and McInroy 2014) illustrates that LGBTQ youth use social media platforms as a tool for actively coming out as they are perceived as safe places. In another study, researchers demonstrate the feasibility of evaluating equality of hospital care for LGBT individuals through analyzing online discussions related to LGBT hospital care on Twitter (Hswen et al. 2018). Haimson (2019) show that the positive sentiment increases over time on average after disclosing their identities to people with close relationships by analyzing posts from Tumblr. Overall, existing research indicates a need to understand better the relationship between the LGBTQ community and their use of social media.
Minority Stress Theory

The minority stress model (Meyer 1995) is a full conceptual framework for understanding how minority stressors impact the mental health of LGBTQ individuals. It defines minority stressors as the chronic stress derived from stigmatized social status in the context of a heterosexist society. Meyer (1995) suggests three types of minority stressors: internalized homophobia, perceived stigma, and prejudice events. Stressors specific to LGBTQ individuals typically contain social stigma, prejudice, and discrimination, such as verbal abuse, harassment, and physical violence.

The minority stress theory has been utilized as a theoretical framework for numerous studies to examine the mental health of gender and sexual minorities. Previous studies have found widespread mental health disparities between the LGBTQ and the broader population in generalized anxiety symptoms (Borgogna et al. 2019), depression (Kulick et al. 2017), and substance abuse (Goldbach, Mereish, and Burgess 2017). Lehavot and Simoni (2011) analyze the links of minority stress with mental health and substance use among women belonging to a sexual minority. Frost, Lehavot, and Meyer (2015) find that minority stress impacts the physical health of the LGBTQ population.

Data

We leverage Twitter timeline data of users who self-disclose as LGBTQ members. The data collection includes four steps: 1) curating a LGBTQ related hashtag list; 2) identifying LGBTQ users; 3) collecting the timeline data of these self-disclosing LGBTQ users; 4) building a control dataset from non-LGBTQ users. We apply the minority stress codebook from (Saha et al. 2019) to annotate tweets in the LGBTQ dataset.

LGBTQ Related Hashtag List

The first step in this process is seeding our sample with two popular LGBTQ relevant hashtags, #gay and #lgbt. We use the Twitter Application Programming Interface (API) to collect the latest 10,000 tweets posted in the US for each seeding hashtag and rank the top 100 co-occurring hashtags by their frequencies. By manually examining the hashtags and removing those with wide-ranging or ambiguous meanings, we find 60 LGBTQ relevant hashtags. After collecting 10,000 tweets containing at least one of the LGBTQ relevant hashtags, we get a corpus of 600,000 LGBTQ relevant tweets.

Twitter Data of Members of the LGBTQ Community

We fetch the bios of the accounts which posted these 600,000 LGBTQ relevant tweets. Bios, are public descriptions of Twitter accounts with a maximum length of 160 characters, where users typically disclose information on who they are. We search these bios using regular expression matching (e.g., gay, lesbian, and bisexual) and pride-related emojis, such as the rainbow flag, pride flag, and transgender symbol. We retrieve 7,241 distinct accounts belonging to self-disclosing LGBTQ Twitter users. From this set, we randomly select 40 users and manually check if their latest 50 tweets explicitly express their identities. Based on manual verification done by the authors, we confirm 32 users to be LGBTQ individuals. The bios of the 8 remaining users contain terms with LGBTQ keywords or emojis, which are either supportive expressions (e.g., LGBTQ life matters) or declare their relationship with the LGBTQ community (e.g., LGBTQ* Advisory Committee in College).

Compiling the LGBTQ Dataset

From the 7,241 unique LGBTQ users, we collect Twitter metadata, including the number of tweets, likes, followers, followees, and the account creation time. Inspired by Pavalanathan and Eisenstein (2016), we filter out untypical users, such as inactive users and advertising accounts by removing users who have more than 5000 followers/followees or who posted fewer than 200 or more than 30,000 tweets. For the remaining 5,708 users, we collect their timeline data from June 1st, 2018 to June 1st, 2021. In the end, the LGBTQ dataset contains 5,708 users. Among these users, 972 users identify as gay, 173 users as lesbians, 1,424 users as transgender, and 95 users as bisexual. For the remaining 3,044 users, they do not identify as part of the LGBTQ community, but are not categorized into one of the former categories.

Compiling the Control Dataset

To compare the linguistic differences between the LGBTQ and the general population during the pandemic, we build a Control dataset of users who do not use LGBTQ keywords in their bios. Recall that our RQ3 is to compare linguistic difference between the general and the LGBTQ population. We seek to isolate the effect of the pandemic between two groups of users who have similar pre-pandemic attributes. To find non-LGBTQ users who have similar linguistic patterns to the LGBTQ ones, we rank the top 100 hashtags from all tweets in the LGBTQ dataset and manually remove hashtags that occur in our LGBTQ related hashtag list, which results in 79 hashtags. For each of these 79 hashtags, we collect 10,000 tweets and the bios of their authors. After removing users whose bios contain LGBTQ keywords and/or LGBTQ emojis, we collect the Twitter metadata of the users and limit the dataset using the same methods as for the LGBTQ dataset. We collect the timeline data of the remaining 36,030 users to build the control dataset.

Annotating Minority Stress on Social media

We use the minority stress codebook based on previous research (Saha et al. 2019), drawn from Meyer’s minority stress framework (Meyer 1995), to annotate minority stress on our LGBTQ dataset. It conceptualizes minority stress as three categories: prejudice events, containing discrimination and violence experiences; stigma, involving expectations of being rejected because of one’s own minority identity; and internalized homophobia, related to the negative attitudes toward oneself. Table II presents some paraphrased tweet examples classified into these categories of minority stress.

To analyze the linguistic differences in minority stress tweets between the time span before and during the COVID-19 pandemic, it is essential to keep the number of tweets in the training dataset balanced between the two time periods. We randomly select 2800 tweets for annotation task, in which 1400 posts are created before the pandemic and 1400...
tweets are created during the pandemic. We refer to them as pre-pandemic dataset and during-pandemic dataset.

**Background of the Annotators** Four authors of this paper and two recruited annotators participate in the annotation process. Four are self-identified females, and two are self-identified males with diverse cultural backgrounds. Besides, one of the annotators belongs to the sexual minority community, which might help in explaining the highly subjective, intrinsic, and complicated context of the tweets to be annotated and the final goals of the annotation task.

**Annotation Task** Four authors independently annotate 200 randomly-sampled tweets and revise the codebook by discussing the discrepancies. Together, the annotators decide to only annotate tweets discussing recent and present experiences pertaining minority stress. Lacking the context of URLs and the complexity of checking the contents in URLs, the URLs in posts are removed but the remaining content of the tweet is included in the dataset.

After establishing the annotation rules, two university students are recruited to annotate the remaining 2,600 tweets. They are asked to learn the codebook with the code explanations, discrepancies, and tweet examples. The two annotators annotate the same sample sets to avoid bias and to increase accuracy. In case annotators have any uncertainties about a tweet, they independently discuss it with the authors. The discussion leads to substantial inter-rater reliability, with an overall Cohen’s kappa (κ) coefficient ([Landis and Koch 1977](https://doi.org/10.1080/00333127708931827)) of 0.727, which is an improvement over the pre-discussion (κ) of 0.542. After discussion, the remaining tweets with no agreement are examined by the annotators and the authors together. Majority voting is applied to decide the final labels. Out of the 2,800 annotated tweets, 684 are categorized as prejudice events, 306 as perceived stigma, and 36 tweets as internalized LGBTQ phobia. Minority stress subcategories are not mutually exclusive and tweets do not provide the full information on the context. So, we subsequently focus on the binary categorization of “minority stress” and “non-minority stress”.

**Methods**

**Study Design and Rationale**

Our study aims to answer three research questions — measuring minority stress using social media posts (RQ1), capturing the linguistic differences in minority stress posts by LGBTQ members before and during the COVID-19 pandemic (RQ2), and contrasting the linguistic difference of LGBTQ members with the general population to explain their emotional reactions to the pandemic (RQ3). For RQ1, we conduct annotations based on the existing taxonomy of minority stress and develop machine learning classifiers using the annotated data. For RQ2, we calculate the coefficient scores for every input feature and analyze the most discriminative features for the best performing models. RQ3, however, involves additional considerations — given that the linguistic difference of LGBTQ members can precipitate due to a variety of reasons beyond their sexual and gender identities, it is essential to contrast these differences with carefully chosen members of the general population. By closely matching a series of pre-pandemic attributes between members of the general population and members of the LGBTQ population, we try to isolate the effect of belonging to a minority group on the respective emotional response to the pandemic. To this end, we adopt a causal inference framework ([Imbens and Rubin 2015](https://doi.org/10.1001/jama.2021.0557)) to address RQ3. The schematic diagram of our approach is shown in Figure 1. Our approach first matches members of the general population (i.e., the control group) with members of the LGBTQ group (i.e., the treatment group) based on pre-pandemic behavioral attributes. For this, we train a machine learning classifier to estimate the likelihood of an individual being assigned to either the treatment or control group (i.e., propensity) based on covariates and perform matching across groups using estimated propensity scores. Within matched groups of control and treatment groups, we analyze the linguistic difference during the pandemic to compare the behavior of similar individuals who differ in terms of their

| Category          | Tweet Example                                                                 |
|-------------------|-------------------------------------------------------------------------------|
| Prejudice Events  | “I was feeling really good until I was called, and then the verbal threats started once they knew I was trans. I’m at a loss for words.” |
| Perceived Stigma  | “I believe that my desire to be visibly gay in games is interfering with my enjoyment of the ones I have. I’m sick of modifications stitching things together and half-measures. I just want to play games and not wear a mask of myself.” |
| Internalized LGBTQphobia | “Being transgender is something I despise. Why couldn’t I just be a cisgirl from the start? Every ray of optimism that you have ends up destroying you. Either you let it kill you for not being yourself, or you let the world kill you for being yourself.” |

Table 1: Sample tweets which belong to different categories of minority stress.
sexual and/or gender identities. In sum, our approach ensures that members of the LGBTQ population and the general population who are being compared have a similar behavior prior to the pandemic. This gives us the means to analyze the linguistic differences of the LGBTQ members, compared to the members of the general population.

**Machine Learning Classifier for Evaluating Minority Stress**

Our methodological approach describes the way that we conceptualize the intrinsic language features within tweets and apply them to a series of machine learning models.

**Feature Engineering** Each tweet in the dataset is described as a vector of the following feature sets:

- **Latent Semantics (Word Embeddings):** To capture language semantics beyond raw keywords, we use word embeddings, which are essentially vector representations of words in latent semantic dimensions. In particular, we use pre-trained word embeddings (GloVe from StanfordNLP [Pennington, Socher, and Manning 2014]) in 50-dimensions that are trained on word-word co-occurrences in a Wikipedia corpus of 6B tokens.

- **Psycholinguistic Attributes (LIWC):** The use of Linguistic Inquiry and Word Count lexicon (Tausczik and Pennebaker 2010) is a well-established approach in extracting psycholinguistic features of texts in multiple dimensions, consisting of affective, cognition and perception, interpersonal focus, temporal references, lexical density and awareness, biological concerns, and social and personal concerns (De Choudhury and De 2014).

- **Sentiment:** Stanford CoreNLP’s deep learning based sentiment analysis tool (Manning et al. 2014) is used to measure the sentiment of the tweets. We use this tool to calculate the positive, negative, and neutral scores for each tweet.

- **Open Vocabulary (n-grams):** Open Vocabulary has been used in a previous study (Kim et al. 2021) to examine the psychological features of individuals. Similar to their approach, we apply top 500 uni- and bi-grams as open vocabulary features in our dataset.

**Machine Learning Models** Using the pre-pandemic and during-pandemic datasets, we build two sets of binary classifiers for detecting tweets with minority stress. We refer to the models trained on the pre-pandemic dataset as **pre-pandemic models** and the ones trained on the during-pandemic dataset as **during-pandemic models**.

Similar to (Saha et al. 2019), we implement and evaluate six models for each training dataset, including a dummy classifier (baseline), naïve Bayes, Logistic Regression, Decision Tree, Support Vector Machine (SVM), and Multilayer Perceptron (MLP) algorithm. For hyperparameter tuning, all classifiers go through k-fold cross validation (k = 10).

**Feature Importance** To understand which features contribute the most to classification, we fit the best performing model to the two training datasets (pre-pandemic and during-pandemic) separately. It turns out that logistic regression models show the best performance. We therefore retrieve coefficient scores for each input variable and compare their importance. In the Results section we focus on the interpretable feature sets, namely LIWC and n-grams. These two feature sets turn out to be the most important feature sets across all feature sets to predict the minority stress. We rank all features with their weights. Features with positive weights contribute to identifying the minority stress tweets, while features with negative weights contribute to the opposite decision of “no minority stress”.

**Matching For Causal Inference**

**Matching Covariates** Our goal is to compare the linguistic difference in tweets between members of the LGBTQ population and the general population during the COVID-19 pandemic. To minimize the possible occurrence of selection bias, matching is an efficient strategy in case-control studies to estimate causal effects. In our case, we build several covariates from the LGBTQ and control datasets using data before December 1st, 2019 to control for similar pre-pandemic behavior on social media. The first set of covariates comprises Twitter users’ social media features (the number of tweets, likes, followers, followees, the account creation time, and posting frequency). The second set contains the distribution of word usage in the Twitter timelines. We extract the top 500 unigrams as the second covariates set. The third set includes the psycholinguistic features in the timeline data by measuring the word distribution in the LIWC lexicon (Wei et al. 2021). As previous research has indicated that mental health is correlated with text readability (Tausczik and Pennebaker 2010), we add several metrics of readability into the last set.

**Propensity Score Analysis** To ensure the similarities between our LGBTQ dataset and the Control dataset, matching is used to pair LGBTQ users and control users, whose covariates are similar to each other. We implement a logistic regression classifier to predict the likelihood of a user belonging to the LGBTQ group or control group based on their covariates. We refer to the likelihood as propensity score.

We divide the propensity scores distribution into 100 strata with equal width. The users with similar propen-
sity scores are grouped into the same stratum (Keciman, Counts, and Gasser 2018). This helps us evaluate the effects of the COVID-19 within each stratum, where the control group users are matched to the LGBTQ users based on the pre-pandemic behavioral traits. We remove the users with propensity scores falling within two standard deviations from the mean. We drop the strata failing to satisfy the minimum sample size within each stratum based on previous causal inference research (De Choudhury et al. 2016). By ensuring that there are at least 50 users per group in each stratum, this approach results in 33 strata, containing 4,107 LGBTQ and 7,129 Control users (Figure 2a).

Quality Assessment of Covariate Matching To assess whether users in the LGBTQ group and control group are statistically comparable, we measured the balance of the covariates. Comparisons between the two groups in each of the 33 valid strata are made by calculating the standardized mean difference (SMD). SMD is the difference in the mean covariate values between the LGBTQ group and the control group, divided by the pooled standard deviation of the two groups. A previous study (Keciman, Counts, and Gasser 2018) has indicated that two groups are balanced if the absolute SMD of all covariates has a value less than 0.2. For the unmatched dataset, the maximum absolute SMD is 0.8, and the mean is 0.25, while the maximum absolute SMD is 0.19 and the mean is 0.07 in the matched dataset (Figure 2b). This satisfies the suggested threshold of less than 0.2.

Estimating the Average Treatment Effects To estimate the effect of the COVID-19 pandemic, we compute the change of word usage in the LIWC variables. The treatment effect (TE) of an outcome variable $S$ is calculated in stratum $k$ by the following equation:

$$TE_k = \frac{\sum_{i} N_k (S_{L_i}^{\text{during}} - S_{L_i}^{\text{before}})}{N_k} - \frac{\sum_{j} M_k (S_{C_j}^{\text{during}} - S_{C_j}^{\text{before}})}{M_k},$$

where $N_k$ is the number of LGBTQ users in stratum $k$, and $M_k$ is the number of matched control users in the same stratum. $S_{L_i}^{\text{before}}$ and $S_{C_j}^{\text{before}}$ are the word usage in the LIWC variables measured for LGBTQ user $L_i$ and control user $C_j$ before the pandemic. $S_{L_i}^{\text{during}}$ and $S_{C_j}^{\text{during}}$ are the word usage in the LIWC variables measured for LGBTQ user $L_i$ and control user $C_j$ during the pandemic.

In this section, we examine if different feature sets affect the performance of the models (RQ1). We investigate the linguistic difference between before and during the pandemic in the context of minority stress (RQ2). We compare the linguistic differences in tweets during the COVID-19 pandemic between the general and the LGBTQ population (RQ3).

**RQ1: Machine Learning Classifier Performance**

We implement and evaluate multiple classifiers in the pre-pandemic training dataset and during-pandemic training dataset, including a dummy classifier (baseline), logistic regression, random forest, support vector machine, naïve bayes, and multilayer perceptron algorithm. Table 2 summarizes the performance of our models and includes the average precision, recall, F1, accuracy, and AUC. We find that logistic regression models show strong and stable performance to indicate minority stress. The best pre-pandemic model shows 75.3% accuracy with an AUC of 0.838, while the best during-pandemic model shows 76.5% accuracy with
AUC of 0.845, which are used for further analysis.

We investigate which feature sets in practice contribute the most to the preference of our classifier. We evaluate our model under five different feature sets: a full feature set as well as each of the four sets excluding one of the features sets. Table 2 summarizes the performance of each feature set in logistic regression models trained on pre-pandemic and during-pandemic datasets. We find that all pre-pandemic and during-pandemic models perform considerably better than the baseline. This reveals that our feature sets do adequately capture the minority stress in tweets, and these different features in the training set do boost the performance of machine learning classification. In addition, we generate the confusion matrix and the receiver operating characteristic (ROC) curves (Figure 2) of the best performing models. We see that the during-pandemic model shows a higher false positive score than false negative score, but the pre-pandemic model has the opposite.

We analyze the misclassified posts to understand what factors contribute to misclassifications. We sample several posts from test sets with the predicted labels by pre-pandemic and during-pandemic models. We find several false positive posts containing the condemnation of discriminatory behavior. For example, “The Supreme Court ruled that companies do not have a right to discriminate against LGBTQ people in the workplace. This historic decision holds firm that LGBTQ people are, and should be, protected from discrimination under federal law.”. This post is annotated as a non-minority stress post, suggesting the advancement of LGBTQ rights. Since there are implicit references of past discrimination behaviors in employment for the LGBTQ population, the models fail to distinguish it from minority stress posts. We notice that some false negative occurs when some posts only state the minority stress but give no details, such as: “It’s sad that we as LGBTQ Americans still cannot help save our own neighbors’ lives! I feel as though it is plain and simple discrimination!”.

RQ2: Linguistic Differences of Minority Stress

To answer RQ2, we examine the most important features for the best minority stress classifiers. We explore which features in tweets contribute to identifying the difference between minority stress in posts before and during COVID-19. These features indicate that the models consider them to be associated with the labels of “minority stress” assigned by human annotators and provide further insights into inferring minority stress from tweets.

Tables 4 and 5 report the top three LIWC features and top five n-gram features, indicating minority stress tweets and non-minority stress tweets with their ranking among all features. MR indicates the maximum rank (607) and ∆ indicates the change of rank with its change direction. For example, feature number with the rank MR in the pre-pandemic model indicates that its rank is 607, which is the most important feature to classify non-minority stress tweets.

In Table 3, we observe that cognitive attributes, including causation, certainty and inhibition, contribute significantly to the classification. Cognitive attribute in LIWC includes words such as “cause”, “know”, and “ought”. Cognition ability is linked to re-structure and understand the emotional experience. Previous studies suggest that an increasing use of cognitive words are linked to health improvement with repeated writing and cognition ability may be essential to adaptation to traumatic or stressful experiences (Klein and Boals 2001). This is also consistent with our annotated dataset. For example, “It took me more than a decade to embrace my homosexuality and that it can never be changed. Even with that knowledge, I am still depressed because sexuality is completely useless to me as an introvert depressed social anxiety sufferer.” Cognitive words, including discrepancy and cause, also contribute to identify the non-minority stress tweets. This might be because some traumatic experiences are caused by their stigmatized identities but other heartbreaking events. Additionally, we find that the during-pandemic model heavily relies on anger, which contains words describing violent actions or angry emotions (e.g., “hate”, “kill”, and “annoyed”). For instance, “I didn’t want others to assume I was gay then try to harm me or figure out where I live for future attack”.

We find many discriminating n-grams that are significant features, such as anti/anti trans, discrimination, discrimination homophobia. Additionally, we also observe that the n-gram lgbtq rights frequently occurs when complaining about Trump’s statements about LGBTQ. For example, “Since Donald Trump took office, his government has...
Table 4: Top three LIWC features for the best performing pre-pandemic and during-pandemic model with its ranking among 607 features. MR indicates the Maximum Rank (= 607). The Δ indicates the change of rank from one model to another with its change direction.

Table 5: Top five n-gram features for the best performing pre-pandemic and during-pandemic model with its ranking among 607 features. MR indicates the Maximum Rank (= 607). The Δ indicates the change of rank from one model to another with its change direction.

RQ3: Linguistic Analysis of Matched Result

To investigate if the COVID-19 pandemic has a statistically significant effect on LGBTQ users, we estimate the effect size (Cohen’s d) in LIWC word usage changes between the LGBTQ group and the control group per stratum. As shown in Figure 24, we find that the average Cohen’s d in the outcome changes across all valid strata is 0.53, which illustrates medium to large differences in LIWC word usages. A Welch’s t-test reveals the statistical significance in these differences \( t \in [-5.64, 3.38], p < 0.05 \), which confirms the significant changes in these outcomes.

We analyze the effect of the pandemic. Figure 25 shows the distribution of the treatment effect across all strata. The average treatment effects greater than 0 indicate a greater increase in the word usage following the impact of the pandemic. We observe that the pandemic causes greater increase for the LGBTQ group on the outcomes of negative \( t = 3.38, p < 0.01 \) and cognitive processes \( t = 3.38, p < 0.01 \). Negative is a category in LIWC containing “no”, “not”, “never”, and so on. One explanation is that because of the disproportionate impact of the pandemic for the LGBTQ groups, LGBTQ users use more cognitive words to express their traumatic experience. We observe the average treatment effect of positive emotion \( t = 1.21, p < 0.05 \) is less than 0, which is translated to worsened observable attribute in the usage of positive words, such as “live”, “nice”, and “sweet”. We find the worse observable attributes of \( t = 1.21, p < 0.05 \), leisure \( t = -4.11, p < 0.01 \) and biological processes \( t = -5.64, p < 0.001 \). The leisure in LIWC contains words related to lifestyles, such as “cook”, “chat”, and “movie”, and the biological processes contains “eat”, “clinic”, “horny”, and so on. Interestingly, there is no significant difference in the usage of negative emotion, which suggests that the pandemic has no disproportionate impact for the LGBTQ group on that.

Conclusion and Discussion

Summary: In this work, we develop theory-driven computational approaches to quantify minority stress from social media data. Our classifiers demonstrate strong capabilities to detect minority stress from tweets. Following this, we investigate the differences in language-related attributes of minority stress before and during the COVID-19 pandemic. We observe the increased prominence of anger words as predictive attributes of minority stress during the pandemic. To explain the differential impact on their emotions, we also contrast the linguistic difference in tweets between LGBTQ population and general population during the pandemic by adopting a propensity score-based matching approach. Our findings reveal that during the pandemic, LGBTQ population are faced with a disproportionate impact on their cognitive and emotional states. This is demonstrated in their wors-
Our findings can provide counselors with the sensitivity and skill to work with diverse backgrounds. According to a study by Meyer (2003), there is a need to expand the framework and distinguish between different stigmatized statuses. Future work could dig deeper into this issue by adding different psychological scales for different groups.

Limitations: Our work is grounded in Meyer’s theory of minority stress. Although it is widely used for various empirical investigations on the health of the minority population, this framework has its own limitations. As acknowledged by Meyer (2003), there is a need to expand this framework and distinguish between different stigmatized statuses. Future work could dig deeper into this issue by adding different psychological scales for different groups.

Another limitation of our study is selection biases. We choose Twitter as our data source as it has many highly active US-based self-disclosed LGBTQ users. This is likely to be influenced by self-selection bias as we can only collect data from those who expressed themselves on social media. This is especially true when considering that LGBTQ people have long been exposed to stigmatization and prejudice.

A further note is that we only use features inside the tweets to measure the minority stress. Other features, such as frequency of posting and author metadata, could be incorporated to analyze the minority stress comprehensively. Another research direction is to explore posts of minority stress across different social media platforms to build accurate, comprehensive, and up-to-date feature sets and gain a better and more thorough understanding of this topic.

Lastly, our analysis does not differentiate results across subcategories of the minority stress due to the difficulty of extracting the context from a single tweet. Future works can analyze the minority stress from the conversation level on Twitter to better understand the minority stress.

Acknowledgements

The calculations presented above were performed using computer resources within the Aalto University School of Science “Science-IT” project.

References

[2016] Andalibi, N.; Haimson, O. L.; De Choudhury, M.; and Forte, A. 2016. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In Proceedings of the 2016 CHI conference on human factors in computing systems, 3906–3918.

[2014] Association, A. C., et al. 2014. American counseling association code of ethics. Alexandria, VA: Author.
[2020] Pietrabissa, G., and Simpson, S. G. 2020. Psychological consequences of social isolation during covid-19 outbreak. *Frontiers in psychology* 2201.

[2019] Saha, K.; Kim, S. C.; Reddy, M. D.; Carter, A. J.; Sharma, E.; Haimson, O. L.; and De Choudhury, M. 2019. The language of lgbtq+ minority stress experiences on social media. *Proceedings of the ACM on human-computer interaction* 3(CSCW):1–22.

[2020] Saha, K.; Torous, J.; Caine, E. D.; and De Choudhury, M. 2020. Psychosocial effects of the covid-19 pandemic: Large-scale quasi-experimental study on social media. *Journal of medical internet research* 22(11):e22600.

[2020] Salerno, J. P.; Williams, N. D.; and Gattamorta, K. A. 2020. Lgbtq populations: Psychologically vulnerable communities in the covid-19 pandemic. *Psychological Trauma: Theory, Research, Practice, and Policy* 12(S1):S239.

[2020] Saud, M.; Mashud, M.; and Ida, R. 2020. Usage of social media during the pandemic: Seeking support and awareness about covid-19 through social media platforms. *Journal of Public Affairs* 20(4):e2417.

[2017] Steinke, J.; Root-Bowman, M.; Estabrook, S.; Levine, D. S.; and Kantor, L. M. 2017. Meeting the needs of sexual and gender minority youth: formative research on potential digital health interventions. *Journal of Adolescent Health* 60(5):541–548.

[2010] Tausczik, Y. R., and Pennebaker, J. W. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology* 29(1):24–54.

[2021] Tsao, S.-F.; Chen, H.; Tisseurasinghe, T.; Yang, Y.; Li, L.; and Butt, Z. A. 2021. What social media told us in the time of covid-19: a scoping review. *The Lancet Digital Health*.

[2022] Verma, G.; Bhardwaj, A.; Aledavood, T.; De Choudhury, M.; and Kumar, S. 2022. Examining the impact of sharing covid-19 misinformation online on mental health. *Scientific Reports* 12(1):8045.

[2020] Wang, Y.; Pan, B.; Liu, Y.; Wilson, A.; Ou, J.; and Chen, R. 2020. Health care and mental health challenges for transgender individuals during the covid-19 pandemic. *The Lancet Diabetes & Endocrinology* 8(7):564–565.

[2021] Wei, J.; Finn, K.; Templeton, E.; Wheatley, T.; and Vosoughi, S. 2021. Linguistic complexity loss in text-based therapy. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4450–4459.

[2021] Xiang, X.; Lu, X.; Halavanau, A.; Xue, J.; Sun, Y.; Lai, P. H. L.; and Wu, Z. 2021. Modern senicide in the face of a pandemic: An examination of public discourse and sentiment about older adults and covid-19 using machine learning. *The Journals of Gerontology: Series B* 76(4):e190–e200.

[2019] Zeeman, L.; Sherriff, N.; Browne, K.; McGlynn, N.; Mirandola, M.; Gios, L.; Davis, R.; Sanchez-Lambert, J.; Aujean, S.; Pinto, N.; et al. 2019. A review of lesbian, gay, bisexual, trans and intersex (lgbti) health and healthcare inequalities. *European journal of public health* 29(5):974–980.

[2021] Zhou, J.; Zogan, H.; Yang, S.; Jameel, S.; Xu, G.; and Chen, F. 2021. Detecting community depression dynamics due to covid-19 pandemic in australia. *IEEE Transactions on Computational Social Systems*. 

