A Study of User Behaviors and Activities on Online Mental Health Communities

by

Nur Shazwani Kamarudin

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved November 2019 by the
Graduate Supervisory Committee:

Huan Liu, Chair
Hasan Davulcu
Arunabha Sen
Xia Hu

ARIZONA STATE UNIVERSITY

December 2019

ABSTRACT

Social media is a medium that contains rich information which has been shared by many users every second every day. This information can be utilized for various outcomes such as understanding user behaviors, learning the effect of social media on a community, and developing a decision-making system based on the information available. With the growing popularity of social networking sites, people can freely express their opinions and feelings which results in a tremendous amount of user-generated data. The rich amount of social media data has opened the path for researchers to study and understand the users' behaviors and mental health conditions. Several studies have shown that social media provides a means to capture an individual state of mind. Given the social media data and related work in this field, this work studies the scope of users' discussion among online mental health communities. In the first part of this dissertation, this work focuses on the role of social media on mental health among sexual abuse community. It employs natural language processing techniques to extract topics of responses, examine how diverse these topics are to answer research questions such as whether responses are limited to emotional support; if not, what other topics are; what the diversity of topics manifests; how online response differs from traditional response found in a physical world. To answer these questions, this work extracts Reddit posts on rape to understand the nature of user responses for this stigmatized topic. In the second part of this dissertation, this work expands to a broader range of online communities. In particular, it investigates the potential roles of social media on mental health among five major communities, i.e., trauma and abuse community, psychosis and anxiety community, compulsive disorders community, coping and therapy community, and mood disorders community. This work studies how people interact with each other in each of these communities and what these online forums provide a resource to users who seek help. To understand users' behaviors, this work extracts Reddit posts on 52 related subcommunities and analyzes the linguistic behavior of each community. Experiments in this dissertation show that Reddit is a good medium for users with mental health issues to find related helpful resources. Another interesting observation is an interesting topic cluster from users' posts which shows that discussion and communication among users help individuals to find proper resources for their problem. Moreover, results show that the anonymity of users in Reddit allows them to have discussions about different topics beyond social support such as financial and religious support.

In dedication to my mother for making me be who I am and my husband for supporting me all the way!

ACKNOWLEDGEMENTS

This dissertation was made possible through the help and support from everyone. However, please allow me to especially dedicate my acknowledgements of gratitude towards the following significant advisors and contributors.

First and foremost, I would like to thank Dr.Liu for his support and encouragement. He has pushed me along throughout this process, supported my various undertakings, and has fully revolutionized my time in graduate school. From the bottom of my heart, thank you Dr.Liu. Secondly, I would like to thank my committee members Dr. Davulcu, Dr. Sen and Dr. Hu for a continuous guidance in helping me finish my research.

I would also thank my senior member in the lab that being helping me throughout my entire journey especially Ghazal. She has shared with me her knowledge, thus making the entire work possible, as well as provided invaluable insight, guidance, and support in a time of dire need. I began this research quite clueless, but as a result of Dr. Liu's and Ghazal's guidance, have not only been able to survive but also thrive. Thank you to all my collaborators that been working with me especially Lydia and Vineeth. With their help and advise, we manage to publish several publications.

Finally, I would like to thank all members of Data Mining Machine Learning Research Lab (DMML) at Arizona State University (ASU) for their constant support and feedback for this work. You guys have been amazing friends to be surrounded with throughout this journey. All the lab meetings, hikings and eating out with you guys will always be the best memories that I ever created in Arizona. You guys have made it possible for me to enjoy this foreign land and survive my entire graduate school.

Last but not least, I am very grateful to have my husband, my mom, my siblings, my entire family and friends who have advised and pushed me forward throughout this endeavor. They have been the biggest support system for me during this roller coaster journey. The product of this dissertation would not be possible without all of them.

This work, in part, is supported by the Ministry of Higher Education Malaysia and University Malaysia Pahang (UMP).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION

For the past few years, social media has become a very popular and well utilized resource for data collection. The main factor that contribute to social media usage is the way people use this medium. Various types of data can be found on social media such as text, image, video, and audio. Social media spreads information faster than any other media. This interaction can be very beneficial if the user receives it in the right way, but at the same time, it can also be toxic for them. It benefits users in different ways such as news alert, traffic message delivery, increased teen awareness, increased marketing exposure, empower individuals to make social change and do social good on a community level. On the other side, social media can also be toxic for users if they use it in different manners. Social media might facilitate cyberbullying, even commit and promote crimes, causes people to spend less time interacting face-to-face, can also correlate with personality and brain disorders, entice people to waste more time, etc. Given these benefits and drawbacks of social media, it has been getting attention from researchers on what's the effect of social media on users as well as how social media can be a source of providing the right information.

A growing body of research has focused on understanding how social media activities can be used to analyze and improve the well-being of people, including mental health (Homan *et al.*, 2014; De Choudhury *et al.*, 2014; Manikonda and De Choudhury, 2017). With the presence of social media data, it is now easier to study the trend of mental health problems and help researchers get information directly from human sources to study mental health issues (Coppersmith *et al.*, 2015a). The easy access to and use of social media allow users to update their social media profiles without time or space restrictions (Park *et al.*, 2012). This makes social media a preferable medium for researchers for their investigations.

Social media has become a good source for data collection, the amount of data on social media data increases rapidly. For example, on Twitter, 350,000 tweets are generated per minute and 500 million tweets are generated per day. With this amount of data, it shows how active people are using social media and data keep on increasing every second. The communication on social media can be either among close friends or between strangers that completely have no idea about each other's identity (or we call it anonymity). This can contribute to a very prolonged interaction than can affect the user's life. For instance, active use of social media with two-way communication can be very beneficial to the user but it can also be destructive or toxic to the user (Robinson *et al.*, 2017).

Social media can affect users in many different ways. The main concern is if social media is beneficial to overcoming mental health problems among users. In News (2017), Facebook was rated as a negative platform when it comes to cyber-bullying and bad sleeping patterns for users. But, when it comes to social support and building online communities, Facebook does help and was rated positively. Hence, it is important to make sure that social media is used in a manner that can benefit all users. Some significant effects of social media contributing to mental health well-being include: (1) social media can reduce stress in users through active communication with other users (Cohan *et al.*, 2017) and also provide information on capturing an individual's present state of mind (Nadeem, 2016); (2) social media is a popular channel for users to seek help and share information on stigmatized issues. The anonymity of social media gives freedom to the user to express their feeling and might also improve in the stigmatized topic discussion (De Choudhury and De, 2014). (3) The use of social media with active communication may lead to an improvement in the capability to share and understand others' feelings (Burke *et al.*, 2010). A study by Grieve *et al.* (2013) indicates that Facebook connectedness may reduce depression and anxiety. Engaging with online communities can also give users the feeling of social appreciation through being understood (Baker and Fortune, 2008).

Content shared on social media platforms has been identified to be valuable in gaining insights into people's mental health experiences (Manikonda and De Choudhury, 2017). Mental health has become a public concern nowadays. People have started to think about the importance of mental health problems and their effects on our society. This is not a minor issue; on the contrary, these problems are very serious issues that can contribute to mental well-being. Users actively share and communicate with online communities and researchers have found that it is a smart idea to leverage social media data to study this problem to help online communities and authorities at the same time. Social media can contribute to an immense change in overcoming this issue.

Figure 1.1 shows different types of mental illnesses and related work using social media data and analysis. Mental illnesses are health conditions involving changes in emotion, thinking or behavior (or combination of these). Existing studies are summarized in categories of mental health issues: (1) mood disorder (Park *et al.*, 2012; De Choudhury *et al.*, 2013; Nadeem, 2016; Reece and Danforth, 2017), (2) post-traumatic stress disorder (PTSD) (Harman and Dredze, 2014; Coppersmith *et al.*, 2015b; Pedersen, 2015), (3) anxiety (De Choudhury and De, 2014; Shickel *et al.*, 2017; El Sherief *et al.*, 2017), (4) psychotic disorder (Mitchell *et al.*, 2015), (5) eating disorder (Chancellor *et al.*, 2016), (6) sexual and gender disorder (El Sherief *et al.*, 2017), (7) suicidal behavior (Kumar *et al.*, 2015; Coppersmith *et al.*, 2015c; De Choudhury and Kiciman, 2017), and (8) attention deficit and hyperactivity disorder (ADHD) (Coppersmith *et al.*, 2015a). Mood disorder such as depression is

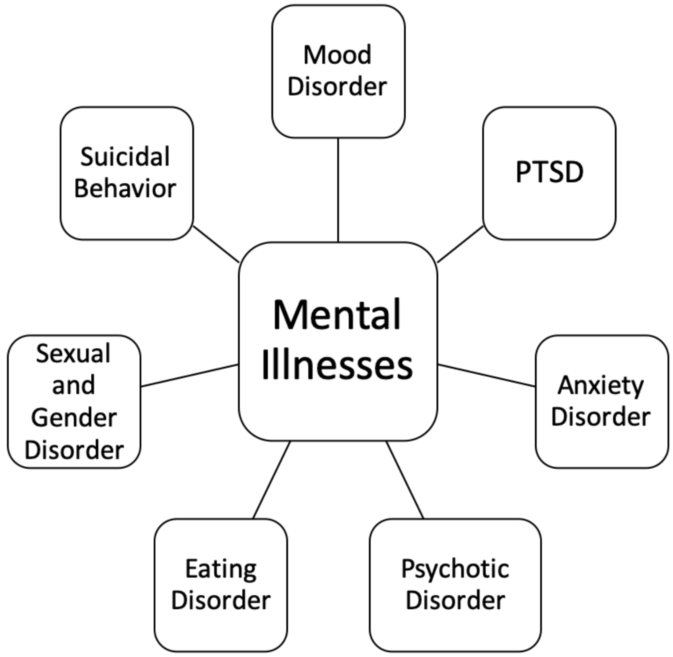


Figure 1.1: An Overview of Different Types of Mental Health Issues and Related Work

studied in (De Choudhury *et al.*, 2014; Coppersmith *et al.*, 2015b; Tsugawa *et al.*, 2015). Depression is one of the mental health issues with high prevalence, receiving increasing attention lately. However, limited work focuses on psychotic disorders, eating disorders, and ADHD. These studies can also leverage social media data as they provide data about individuals' language and behavior (Harman and Dredze, 2014). Researchers use social media data to predict types of mental issues (De Choudhury *et al.*, 2013, 2014; Kim *et al.*, 2016; Chancellor *et al.*, 2016). In Nadeem (2016), the severity of users' mental issues is estimated using social media data. A similar finding is reported by Mitchell *et al.* (2015). Network information available in social media data is also leveraged for studying mental health issues (Kawachi and Berkman, 2001; De Choudhury *et al.*, 2013). Given all these works, it motivates us to carry on with our work by studying user behaviors using social media data. We have two pieces of works that we compiled together in this dissertation. The first work was focusing on mental health in the scope of sexual abuse. For the second work, we decide to study on broader mental health communities by studying various online mental health subreddits which are available on Reddit.

This dissertation studies user behavior and user activities on the online mental health communities. First, we study a variety of user behavior in the scope of sexual abuse. We investigate the role of such online forums in delivering useful information to people seeking psychological and mental support. By leveraging natural language processing (NLP) techniques on rape-related posts from Reddit, we unravel the nature and personality of both victims and respondents. Specifically, we explore the following topics: (a) analyze the diversity of topics that emerge from posts and replies (i.e., subreddits), (b) understand the concerns and obstacles faced by victims, (c) examine the nature of responses to see if their scope extends beyond emotional support, and (d) understand how online responses differ from the real-world responses.

Second, we extend the same scope of work on comparing two subreddits that related to rape and sexual abuse which are r/rape and r/rapecounseling. These two subreddits' discussion content are about "rape", but /r/rapecounseling subreddit involves professionals and /r/rape subreddit does not. Given this content knowledge, this study points out the differences between these subreddits and presents interesting findings. We employ natural language processing techniques to extract topics of responses, examine how diverse these topics are to answer research questions such as whether responses are limited to emotional support; if not, what other topics are; what the diversity of topics manifests.

Finally, we extend the work on multiple mental health communities on Reddit. In particular, we investigate the potential roles of social media on mental health among five major communities, i.e., trauma and abuse community, psychosis and anxiety community, compulsive disorders community, coping and therapy community, and mood disorders community. We study how people interact with each other in each of these communities and what these online forums provide a resource to users who seek help. To understand users' behaviors, we extract Reddit posts on 52 related subcommunities and analyzes the linguistic behavior of each community. We report linguistic analysis to understand each mental health community. We first study the sentiment per each Reddit community and follow by analyzing the topics. By utilizing all related posts from each subreddits in each community, we establish the following exciting insights: (a) We observe that the headline of a post does not indicate the whole content of that discussion post, (b) we found that sentiment distribution across mental health communities was higher on positive sentiments as compared to negative sentiments, and (c) topic distributions for each community vary but there exist similarities between them.

This dissertation is organized into six chapters. Chapter 2 discusses social media data on mental health and related work. Chapter 3 presents a set of approaches that we used in this research. Next, our work which analyzes mental health in the scope of sexual abuse is discussed in Chapter 4. After that, Chapter 5 explains our extensive work on user activities on online mental health communities. Lastly, the study is concluded and a list of possible future works is presented in Chapter 6.

Chapter 2

RELATED WORK

Social media has become a good source for data collection. There are different types of data that can be used from social media such as text, image, video, and audio. The amount of data on social media data increases rapidly. For example, on Twitter, 350,000 tweets are generated per minute and 500 million tweets are generated per day. A major factor that might affect social media users is the way they use social media because it can be very beneficial or toxic at the same time.

2.1 Social Media Data on Mental Health

Various types of social media data have been used by researchers to study mental health. Existing work could be categorized into three groups based on the type of utilized social media data, 1) linguistic-based data, 2) visual-based data, and 3) combination of linguistic and visual data. Next, we will discuss each separately.

2.1.1 Linguistic-based Data

Over the past few years, research in crisis informatics has utilized language as a medium to understand how major crisis events unfold in affected populations, and how they are covered in traditional media as well as online media such as blogs and social media sites (Saha and De Choudhury, 2017). Studies have shown that social media can provide a comforting environment for support seekers especially when it comes to stigmatized issues that make them reluctant to share with individuals around them.

Social media has been accordingly used to understand users' mental health issues. Interesting work by Coppersmith *et al.* (2015c) studies suicide attempts or even ideation among users using their posts on Twitter. The authors crawl tweets from 30 geolocations from all over the United States with at least 100 tweets per location. Then, they use natural language processing techniques to compare the behavior of users who attempted suicide with those users who have previously stated that they were diagnosed with depression and neurotypical controls. In another work, Park *et al.* (2012) focus on studying the impact of online social networks (OSN) on the depression issue. Accordingly, they collect two different kinds of data: 1) data from Internet-based screening test, which includes information of 69 participants who were asked to complete a questionnaire including depression related questions, and 2) collected tweets from Twitter from June 2009 to July 2009 that include the keyword 'depression'. After qualitative and quantitative analysis of collected data, authors show that social media data could be used to understand users' mental health issues.

Nadeem (2016) use social media data to study Major Depressive Disorder (MDD) issues in individuals. They use a publicly available dataset that was built from Shared Task organizers of Computational Linguistics and Clinical Psychology (CLPsych 2015) (Coppersmith *et al.*, 2015b). This dataset includes information on Twitter users who were diagnosed with depression. In another work, Amir *et al.* (2017) use Twitter to study depression and post-traumatic stress disorder (PTSD). They investigate the correlation between users' posts and their mental state. In particular, they investigate if tweets could be used to predict whether a user is affected by depression and PTSD or not. De Choudhury *et al.* (2013) also show that social media can be utilized for predicting another mental health condition, i.e., major depressive disorder (MDD), using Twitter data. In this work, authors employ crowd-sourcing techniques to provide the ground truth for their experiments. In another work by El Sherief *et al.*

(2017), authors use Twitter data to study gender-based violence (GBV). Purohit *et al.* (2015) use Twitter's streaming API to sample a set of GBV related tweets based on a set of key phrases defined by the United Nations Population Fund (UNFPA).

Another well-known social media platform is Reddit, which is a forum-based social media platform that captures the communication between the original post and the user who left a comment on the thread. Each thread discusses a specific topic, which is known as a "subreddit". The work by De Choudhury and De (2014) uses Reddit to investigate how users seek mental health related information on online forums. They crawl mental health subreddits using Reddit's official API [1] and Python wrapper PRAW [2] . In another work, De Choudhury and Kiciman (2017) use Reddit to study the language style of comments left by users on the discussion forum in terms of influences towards suicidal ideation. This work fills the gap of how online social support can contribute to this specific problem. Authors use stratified propensity score analysis to determine if the user was affected by comments or not. They also estimate the likelihood that a user will receive a treatment based on the user's covariates. There was a study by De Choudhury *et al.* (2016) with work specifically on mental health subreddits such as r/depression, r/mentalhealth, r/bipolarreddit, r/ptsd, r/psychoticreddit. Based on the time stamp, data were divided into a treatment group and control group. These groups were further divided based on the causal analysis to analyze the effect of comments on the content of earlier posts with comments shared and received by users in their dataset.

Another work by Saha and De Choudhury (2017) uses Reddit data to study the effect of gun violence on college students and the way they express their experience on social media. Pan and Yang (2010) collect related data from Reddit. Then they

[1]http://www.reddit.com/dev/api

[2]https://praw.readthedocs.org/en/ latest/index.html

develop an inductive transfer learning approach to see the pattern of stress expression by computing the mean accuracy value. In particular, they first build a classifier that labels the expressed stress in posts as *High Stress* and *Low Stress*. Then, they adopt the trained classifier to categorize collected posts from Reddit to identify posts that express a higher stress level after the shooting incident. Another work from Lin *et al.* (2017) focuses on how online communities can affect the development of interaction within social media. This work investigates if new members will bring interruption in terms of perspectives towards the social dynamic and lower content quality. The authors accordingly generate three questions related to user reception, discussion content and interaction patterns. They use Reddit data from Google BigQuery by choosing the top 10 subreddits between the years 2013 and 2014. They study user reception, post content, and commenting patterns among Reddit users. This work studies the role of online social support based on historical data in conjunction with its effect on future health. It also investigates linguistic changes in online communities over time using data from two peer-reviewing communities. Cohan *et al.* (2017) study mental health by analyzing the content of forum posts based on the sign of self-harm thoughts. Their main goal was to study the impact of online forums on self-harm ideation. They consider 4 levels of severity for the post content. The author builds a model that includes lexical, psycholinguistic, contextual and topic modeling features. Their data were collected from a well-known mental health forum in Australia, namely, ReachOut.com [3] .

Farnham and Churchill (2011) works on setting the boundaries between personal and social technologies to catch the also an interesting topic to be discussed. It is not easy to get a direct response using social media data. So authors run this experiment to find out the answer to their question related to this issue. Facet is multiple sides

[3] https://au.reachout.com/

of the identities of an individual that might change based on social situations. To study this issue, the author distributes a questionnaire investigating how people facet their identities and their lives. Then, they study how these groups of people use email and Facebook to express themselves based on their facets. The author designs an online questionnaire and distributed through Yahoo! networks within two weeks' timestamps. They were able to get 631 participants of various ages (18 years old and older).

2.1.2 Visual-based Data

The popularity of visual-based social media has increased rapidly. Users tend to communicate on social media by posting their photographs. Photo sharing provides a unique lens for understanding how people curate and express different dimensions of their personalities (Andalibi *et al.*, 2015). People use photos to define and record their identity, maintain relationships, curate and cultivate self-representation, and express themselves (Van House and Davis, 2005).

Posting pictures have become one way of communication among social media users. The definition of "selfie" is a picture that users take of themselves. Kim *et al.* (2016) study the behavior of selfie-posting using Instagram data. This work uses selfie-posting to predict the intention of users who post selfies on social networking sites (SNSs). This work defines five hypotheses before they start designing the experiment. Those hypotheses were: attitude toward the behavior of selfie-posting, subjective norm, perceived behavioral control, and narcissism, which were possibly related to the intention to post selfies on social network sites (SNSs). The last hypothesis was the intention to post selfies on SNSs is positively related to the actual selfie-posting behavior on SNSs. They begin with 89 Instagram users. They recruit these users based on their agreement to be part of the study. Two coders analyze each user's account and the total sample size was ($n = 85$). From the total number of participants, 9 were males and 76 were females. They also count the total number of the pictures posted on each user's account in a 6-week timestamp. Each user was required to answer a list of questions that were related to the standard Theory of Planned Behavior (TPB) variables such i.e., attitude, subjective norm, perceived behavioral control, and future intention based on (Ajzen, 1991).

Another work by Reece and Danforth (2017) uses visual data for studying mental health in social media. The authors use Instagram data from 166 individuals with 43,950 photographs. To study the markers of depression, they use machine learning tools to categorize users into healthy and depressed groups. They began their experiment by crawling all posts from each user's account upon their agreement. Participant users were also required to answer a depression-related questionnaire that contained specific questions based on inclusion criteria. In the last step of experiments crawled Instagram photos are rated using a crowd-sourced service offered by Amazon's Mechanical Turk (AMT) workers.

2.1.3 Combined Data of Linguistic and Visual-based

Apart from using only visual data or only linguistic data, researchers also combine these two kinds of data to study social media influence on mental health. Sociologists also claim that it is not possible to communicate by using the only words; people also use pictures to communicate with each other (Bourgeault *et al.*, 2010). A work by Burke *et al.* (2010) used different features in Facebook data such as wall posts, comments, "likes", and consumption of friends' content, including status updates, photos, and friends' conversations with other friends to study the role of directed interaction between pairs. This work distinguishes between two types of activity: directed communication and consumption.

To do this, they recruited 1199 English-speaking adults from Facebook to be their research participants. Andalibi *et al.* (2015) study depression-related images from Instagram. They use image data and the matching captions to analyze if a user was having a depression problem or had faced this kind of problem in the past. They were keen on investigating if this group of users engage in a support network or not, and how social computing could be used to encourage this kind of support interaction among users. They gather 95,046 depression tagged photos posted by 24,920 unique users over one month (July 2014) using Instagram's API. All public details of each image were stored from these photos, such as user ID, number of likes and comments, date/time of creation, and tags. After conducting data collection, they begin the experiment by analyzing images and their textual captions. They also develop a codebook that includes 100 sample images and captions. Those coders then manually discuss the codebook to provide the best result for the experiment. Then, they add 100 more sample images and repeat the same steps.

Likewise, Peng *et al.* (2005) investigate the effect of pets, relationship status, and having children, towards user happiness using Instagram pictures and captions. They use several hashtags such as #mydog, #mypuppy, #mydoggie, and #mycat, #mykitten, #mykitty to gather images of pet owner from Instagram. For non-pet owners, hashtags such as #selfies, #me, and #life were used to crawl the data. Before they started with the experimental steps, they began by classifying their data. The authors also provided the processed human face data called the face library for other researchers to use (see Table 2.1).

Manikonda and De Choudhury (2017) use popular image-based media data in order to study mental health disclosure. This work extracts three main visual features from each image that they have in the corpus. Those features include visual features (e.g., color), themes, and emotions. Authors crawl the data from Instagram

Table 2.1: List of Mental Health Related Social Media Data and Available Datasets.

Type of Data	Paper	Name	Availability
Linguistic data	(Nadeem, 2016; Coppersmith *et al.*, 2015c)	Computational Linguistic and Clinical Psychology (CLPsych)	Publicly available at https://bit.ly/2T0hMGO
	(Saha and De Choudhury, 2017; Lin *et al.*, 2017)	Google News dataset	Publicly available at https://bit.ly/1LHe5gU
	(Park *et al.*, 2012; Nadeem, 2016; El Sherief *et al.*, 2017; De Choudhury and De, 2014)	Reddit, Twitter, Facebook	Crawled using provided API by social media platform
Image data	(Kim *et al.*, 2016)	Instagram	Publicly available at https://bit.ly/2SV5cbY
Combination of text and image data	(Andalibi *et al.*, 2015; Manikonda and De Choudhury, 2017)	Twitter, Facebook, Instagram	Crawled using provided API by social media platform
Link data	(De Choudhury *et al.*, 2013; Kawachi and Berkman, 2001)	Twitter, Facebook	Crawled using provided API by social media platform

using Instagram's official API [4] . The main focus of this work was to study mental health disclosure based on visual features, emotional expression and how visual themes contrast with the language in a social media post. They specifically choose 10 mental health challenges from Instagram before they crawl two million public images and textual data from that particular medium. Those categories contain 10 types of disorders, which were: anxiety disorder, bipolar disorder, eating disorder, non-suicidal self-injury, depressive disorder, panic disorder, OCD, PTSD, suicide, and schizophrenia. Before they begin their experiment, they consult with the Diagnostic and Statistical Manual of Mental Health Disorders (DSM-V) in order to confirm that their final disorder categories could be reliable.

To sum-up, user-generated social media data is heterogeneous and consists of different aspects such as text, image, and link data. Table 2.1 summarizes different datasets used by researchers to study mental health using social media.

2.2 Studying Mental Health in Social Media

As social media data has recently emerged as the main medium to spread information among online communities, there are also various approaches used by researchers to study-related problems. This section discusses the approaches for studying mental health on social media. In this section, we elaborate on the types of techniques or

[4] https://www.instagram.com/developer/

tools used in their research. Figure 2 shows the categorization of social media analysis, namely machine learning methods, feature engineering, and survey methods. Next, we introduce how social media analysis is used for mental health analysis in social media.

2.2.1 Machine Learning Methods

We discuss machine learning methods in terms of classification, clustering, and prediction with social media data for studying mental health problems.

Classification

In order to estimate the likelihood of having depression among users within a dataset, a work by Nadeem (2016) employ four types of classifiers (Decision Trees, Linear Support Vector Classifier, Logistic Regression, and Naive Bayes). They present a set of attributes to characterize the behavioral and linguistic differences of two classes. To do that, author utilize scikit-learn [5] which is a popular tool with many supervised and unsupervised machine learning algorithms (Pedregosa *et al.*, 2011).

Clustering

De Choudhury *et al.* (2013) cluster the ego-networks among users on social media. By clustering the ego-network, they study the characteristics of the graphs based on the egocentric measures, such as the number of followers, number of followees, reciprocity, prestige ratio, graph density, clustering coefficient, 2-hop neighborhood, embeddedness and number of ego components. Another work by Kawachi and Berkman (2001) studies the correlation of social ties and mental health and finds that depressed individuals tend to cluster together.

[5] http://scikit-learn.org/stable/

Prediction

The work by Reece and Danforth (2017) predicts depression using photographic details, such as color analysis, metadata components and algorithmic face detection. In another approach, De Choudhury *et al.* (2013) divide users into two categories based on differences in behavior. For each user, they utilize a set of behavioral measures, such as mean frequency, variance, mean momentum, and entropy based on a user's one-year Twitter history. In order to avoid overfitting, the authors employ principal component analysis (PCA), then compared their method with several different parametric and non-parametric classifiers.

2.3 Evaluation Methods

In this section we discuss the utilized evaluation metrics and findings from the aforementioned papers. Figure 2.1 represents the utilized evaluation metrics for a mental health studies in social media. We begin this section by overviewing evaluation metrics. Then, we discuss the findings from previous works.

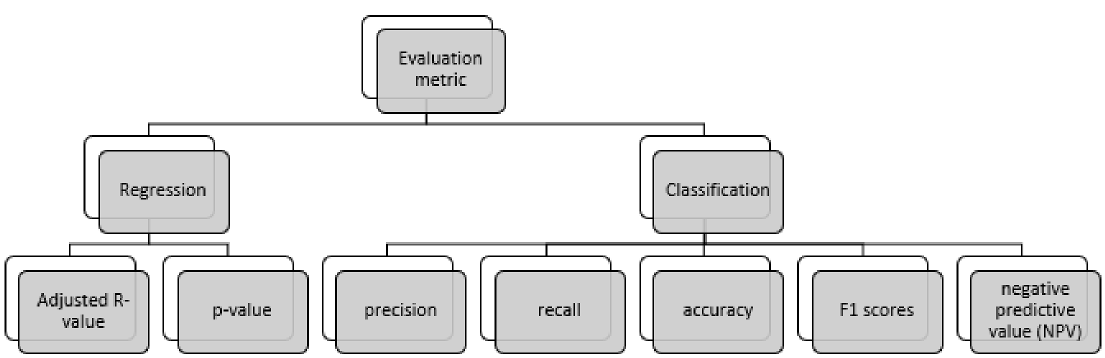


Figure 2.1: Categories of Evaluation Metrics Used for Mental Health Analysis in Social Media.

2.3.1 Evaluation Metrics

There are various evaluation metrics in data mining analysis. We discuss the evaluation metrics used for studying mental health using social media data. The most common prediction metrics include precision, recall, F1 scores, and recipient operating classification (ROC) curves. Equation 1 shows the calculation of precision defined as the number of true positives (TP) divided by the sum of all positive predictions, TP and false positive (FP). Equation 2 is defined as the number of TP divided by the sum of all positives in the set, TP and false negative (FN). Equation 3 measures F1 scores by considering both precision and recall. The F1 score is the harmonic average of the precision and recall, and an F1 score reaches its best value at 1. Another metric known as adjusted R-squared is represented in equation 4. Adjusted R-squared is often used for explanatory purposes and explains how well the selected independent variable(s) explain the variability in the dependent variable(s). In adjusted R-squared, n is the total number of observations and k is the number of predictors. Adjusted R-squared is always less than or equal to R-squared.

$$precision = \frac{TP}{TP + FP} \tag{2.1}$$

$$recall = \frac{TP}{TP + FN} \tag{2.2}$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{2.3}$$

$$(R^2_{adj}) = 1 - [\frac{(1 - R^2)(n - 1)}{n - k - 1}] \tag{2.4}$$

Several works implement these metrics for their experiments (De Choudhury *et al.*, 2013; Nadeem, 2016; Cohan *et al.*, 2017; Saha and De Choudhury, 2017). De Choudhury *et al.* (2013) evaluate their proposed classification approach by predicting individual depression level from their posts. They use precision, recall, accuracy, and receiver-operator characteristic (ROC) for evaluation. Their experimental results show that their classifier has a good performance in depression prediction. Another work from De Choudhury and Kiciman (2017) measures the most positive or negative z scores in order to differentiate between mental health with influence risks to suicidal ideation (SW) users and mental health users. On the contrary, a study by Park *et al.* (2012) uses the coefficients from regression models to predict the Center for Epidemiologic Studies Depression Scale (CES-D) score. It then evaluates the proposed approach by measuring adjusted R-squared (equation 4) and p-value. Another work by Amir *et al.* (2017) measures *F1* and *binary F1* with respect to a mental condition in order to measure the performance of different models for its experiment.

Furthermore, El Sherief *et al.* (2017) measure the favorite rate and retweet rate for each tweet in order to count how many times a tweet was favorited and retweeted, respectively. These metrics are used to explore the engagement of users with gender-based violence (GBV) content on Twitter. Saha and De Choudhury (2017) measure accuracy, precision, recall, F1-scores and ROC-AUC in order to see the performance level of their stress predictor classifier. Likewise, Coppersmith *et al.* (2015c) plot ROC curve of the performance for distinguishing people who attempted suicide from their age- and gender-matched controls. In order to compare the accuracy of all data and pre-diagnosis in their model prediction, Reece and Danforth (2017) measure recall, specificity, precision, negative predictive value (NPV) and F1-scores.

Furthermore, a study by Manikonda and De Choudhury (2017) calculates the Spearman rank correlation coefficients to compare the most frequent tags across all pairs of visual themes that belonged to six visual themes of mental health-related posts. Burke et al. present the ordinary least square (OLS) regressions for bridging and bonding social capital and loneliness based on the overall SNS activities (Burke *et al.*, 2010). Additionally, Lin *et al.* (2017) study the effect of newcomers to existing online forums such as Reddit. They leverage the regression analysis by calculating the adjusted R-squared in order to measure the average score of post voting and complaint comment percentage on the content of the subreddit. In another work, Andalibi *et al.* (2015) calculate Cohen's Kappa coefficient to analyze depression related images along with the textual captions.

Chapter 3

AN NLP APPROACH TO MENTAL HEALTH STUDY

Over the past few years, research in crisis informatics have utilized language as a medium to understand how major crisis events unfold in affected populations, and how they are covered on traditional media as well as online media such as blogs and social media sites (Saha and De Choudhury, 2017). Studies have shown that social media can provide a comforting environment for support seekers especially when it comes to stigmatized issues that make them reluctant to share with individuals around them.

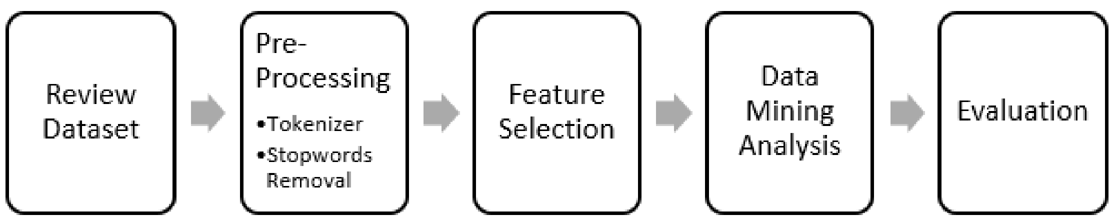


Figure 3.1: Steps of Social Media Analysis.

Figure 3.1 shows the common steps in social media analysis. It starts with a dataset review in which the researchers need to choose the right dataset for their experiment. The second step is data pre-processing, which means preparing the data for the experiment such as removing stop words or word/sentence tokenizing. The next step is to select meaningful features from social media data such as an image or textual features. After selecting the right features, it is a data mining analysis that includes deploying various techniques to develop the desired model.

The final step is an evaluation, employing different metrics such as accuracy, recall, precision, F1 scores, for example. Natural language processing plays a very important role in linguistic social media analysis. This chapter discusses feature representation techniques used in studying social media for mental health in this dissertation.

3.1 N-grams

This text representational technique is widely adopted and is basically a set of co-occurring words within a given window. Features extracted using this technique are based on word frequency counts. N-gram model predicts the occurrence of a word based on the occurrence of its N - 1 previous words. In De Choudhury and De (2014), authors calculate the most frequent unigrams from all Reddit posts and use negative binomial regression as their prediction model.

Saha and De Choudhury (2017) built a supervised machine learning model in order to classify stress expression in social media posts into binary labels of *High Stress* and *Low Stress*. To develop the transfer learning framework for their experiment, they measure the linguistic equivalence by borrowing a technique from domain adaptation literature (Daume III, 2009). By using top n-grams ($n = 3$) as an additional feature, Saha et al. developed a binary Support Vector Machine (SVM) classifier to detect *High Stress* and *Low Stress*. To build their training set, the authors extract 500 n-grams from the Reddit posts that they crawled. They compute the cosine similarity and compare their data with a Google News dataset in a 300-dimensional vector space. Authors find that it is possible to use social media content to detect psychological stress. On the other hand, as Twitter data has a limited number of characters per post, another work by Braithwaite *et al.* (2016) designs a character n-gram language model (CLM) to get the score for each short text. This specific method examines sequences of characters including spaces, punctuation, and emoticons. For example, if we have a set of data from two classes, the model is trained by recognizing the sequence of characters. Similar character sequences will be classified into the same class. Given a novel text, the model can do estimations on which class can produce and generate all the texts.

Furthermore, Manikonda and De Choudhury (2017) extract n-grams (n=3) to check the suitability and reliability of their corpus. Extracted n-grams have been further used to investigate if they are facing mental health disclosures. In order to extract visual features from the dataset, the authors pair OpenCV and Speeded Up Robust Features (SURF) (Bay *et al.*, 2006). This approach is able to identify the meaningful themes from images. To study the linguistic emotions based on the visual themes, this work uses psycholinguistic lexicon LIWC and TwitterLDA. These two approaches help authors to measure the estimation of how themes and images were coherent to each other.

3.2 Linguistic Inquiry and Word Count (LIWC)

It is a text analysis application that can be used to extract emotional attributes on mental health. This tool will be able to extract psycholinguistic features (Cohan *et al.*, 2017). Manikonda and De Choudhury (2017) use LIWC [1] on texts associated with mental health images spanning different visual themes. LIWC can also characterize linguistic styles in posts from users (Rude *et al.*, 2004). Park *et al.* (2012), use LIWC to quantify the level of depressive moods from their Twitter data. They compare a normal group vs. depressed group by measuring the average sentiment score from categories provided by the tools. LIWC contains a dictionary of several thousand words and each word fed to this tool will be scaled across six predefined categories: *social*, *affective*, *cognitive*, *perceptual*, *biological processes*, and *relativity*.

[1] http://www.liwc.net/

Every criterion will have its own categories and sub-categories. For each sub-category, LIWC will assign specific scores for each word. De Choudhury *et al.* (2013) use LIWC to study Twitter users' emotional states. Then, they use point wise mutual information (PMI) and log-likelihood ratio (LLR) to extract more features from their corpus. El Sherief *et al.* (2017) leverage LIWC in order to measure interpersonal awareness among users by differentiating perceived user and actual user characteristics.

In addition, De Choudhury and De (2014) captures the linguistic attributes of their data by measuring the unigram and then employ psycholinguistic lexicon LIWC. They choose LIWC because it can categorize Redditors' emotions. They also examine the factors that drive social support on mental health Reddit communities, where the authors build a statistical model by measuring the top most frequent semantic categories from LIWC. The authors in De Choudhury and Kiciman (2017) adopt the LIWC lexicon to study the various sociolinguistic features from their dataset and then measure the t-tests in order to analyze the differences between subpopulations. Coppersmith *et al.* (2015c) also use LIWC to study the pattern of language in conjunction with psychological categories generated from their dataset. This work uses LIWC to interpret how language from a given psychological category will be scored by the classifiers that they built. Likewise, Saha and De Choudhury (2017) investigate on quantifying the psycholinguistic characterization. They employ LIWC measures to understand the psychological attributes in social media.

3.3 Topic Modeling

One of the most mentioned topic modeling methods is Latent Dirichlet Allocation (LDA), which works by drawing distribution topics for each word in the document (Blei *et al.*, 2003). Then, words are grouped based on the distribution value. Similar words are in the same topic category. Cohan *et al.* (2017) use the LDA model to find a set of topics from their data collection. By training the LDA topic model on the entire forum posts from their dataset, they are able to use the topic model as additional features for their experiment, which boosted the performance of their system and prove the effectiveness of topic modeling. Additionally, Manikonda and De Choudhury (2017) use TwitterLDA to extract the linguistic themes from their dataset to see if visual and text are coherent to each other when it comes to mental health disclosure on Instagram.

Amir *et al.* (2017) adopt a model known as Non-Linear Subspace Embedding (NLSE) approach that can quantify user embedding based on Twitter post histories (Astudillo *et al.*, 2015). The authors evaluate user embedding by using User2Vec (u2v), Paragraph2vec's PV-dm and PV-dbow models. They also leverage Skip-Gram in order to build vectors. Another design based on bag-of-embedding was bag-of-topics by using LDA to indicate topics presented in the user's posts. De Choudhury and De (2014) leverage LDA to identify types of social support on Reddit. They also consider information on practices that people share with the communities by characterizing self-disclosure in mental illness. The authors find that Reddit users discuss diverse topics. These discussions can be as simple as talking about daily routines but it can also turn into a serious discussion that involves queries on diagnosis and treatments.

Additionally, a work by Lin *et al.* (2017) studies linguistic changes and for their data, the authors use several post-level measures including cross-entropy of posts and Jaccard self-similarity between adjacent posts. Then, the authors use the LDA model in order to compare the topic distribution among posts and general Reddit post samples. They also track the linguistic changes in sub-communities. In order to examine the interaction network's structural change, they calculate the exponent α in the network's power-law degree distribution which gives the graph densification of the network (Chancellor *et al.*, 2016). Reddit allow users to vote on each post and comments, and the authors leverage this feature by computing the average score and complaint comment percentage in order to investigate community reaction to the content produced by newcomers.

Chapter 4

MENTAL HEALTH IN THE SCOPE OF SEXUAL ABUSE

Online forums such as Reddit and Volt have emerged as a portal for victims and survivors of rape and sexual abuse to come together in a mutually supportive and safe environment (Andalibi *et al.*, 2016). Rape is a subject of social stigma that restricts victims from openly sharing their experience due to the following reasons: (1) concern about confidentiality, (2) fear of not being believed and (3) feeling of guilt and embarrassment (Ullman and Filipas, 2001; Sable *et al.*, 2006). Unlike traditional support groups, the anonymity of online forums encourages users to break the mold and express their emotions and feelings with less fear and guilt.

Analyzing social forums for understanding stigmatized issues is a relatively new field of research. For example, a study by Ma *et al.* (2016) show the degree of self-disclosure and support seeking on social media. In another work, Cohan *et al.* (2017) use statistical tests to understand the correlation between social support and suicidal ideation. De Choudhury and De (2014) also explore the problems of self-disclosure from a perspective of mental health. Despite providing insightful outcomes, these studies fail to understand the linguistics facets of user responses and their usefulness to the main issue. Additionally, they rely heavily on human-computer interaction models to validate their hypothesis; unfortunately, this is a costly and time-consuming process. To address these shortcomings, in this chapter, we extract Reddit posts from two related subreddits which are /r/rape and /r/rapecounseling. Our work was designed to understand the nature of user responses by addressing the following questions:

(RQ1) What are the most frequently discussed topics?

(RQ2) Are responses always emotional and sympathetic?

(RQ3) What are the differences between the selected subreddits in terms of emotion and discussion topic? Which community shows better score in term of sentiment analysis?

(RQ4) What are other topics that are found in response to a victim's query?

(RQ5) What are the ramifications of these diverse topics?

We answer these questions using a variety of linguistic and natural language processing (NLP) techniques. First, we apply n-gram model to understand the frequency and the nature of the language used in the discussions. Second, we use psycholinguistic tools, i.e. LIWC (Linguistic Inquiry and Word Count), to show that user responses are not limited to emotional support. Then, we examine whether the extracted emotions of the two subreddits are statistically different from each other or not. Third, we apply topic modeling to provide meaningful responses that are relevant to the posts of victims. Finally, we conduct sentiment analysis in order to study the relationship between the emotions of people from each chosen subreddits by analyzing the post's conversation content.

4.1 Data Collection

Although there are several online forums, Reddit [1] has emerged as the best candidate for exploring stigmatized issues (Sable *et al.*, 2006; Manikonda *et al.*, 2018). For example, Andalibi *et al.* (2016) use Reddit to find linguistic differences between

[1] www.reddit.com

temporary and anonymous accounts (also termed as *throwaway* accounts) and identified accounts in the context of sexual abuse to find that men are most likely to use throwaway accounts when they post about sexual abuse. Sable *et al.* (2006) show that semi-anonymity of Reddit enables candid self-disclosure around stigmatized topics such as mental health and Ullman and Filipas (2001) use Reddit to investigate health-related information seeking. Since these studies have clearly demonstrated the potential benefits of the Reddit forum, we decided to use this platform for crawling and compiling our dataset. We crawled the set of all posts that correspond to the topics on [/r/rape [2] and /r/rapecounseling [3]] [4] . These two subreddits discussion content are about "rape" but /r/rapecounseling subreddit involve professionals and /r/rape subreddit does not. Given this feature, this study points out the differences between these subreddits and found interesting finding as explained further in the upcoming section. Reddit platform is structured based on conversational threads where users post their query and other users respond to these queries in the form of support, feedback, and suggestions (De Choudhury and De, 2014). From every post, we collect a variety of meta-data such as the title of the post, textual contents, timestamp (i.e., the time of posting), author id, and the number of upvotes and downvotes. Our first dataset consists of 1921 unique posts from the combination of both subreddits, with an average of ten responses per post. We employed several text preprocessing techniques in our data collection to reduce noise and prepare data for the analysis.

[2] https://www.reddit.com/r/rape/

[3] https://www.reddit.com/r/rapecounseling/

[4] The collected data is publicly available at https://github.com/nkamaru1/RedditData

4.2 Understanding Common Problems

We apply a variety of natural language processing (NLP) techniques to extract meaningful features that address our questions. Our experimental methodology involves quantitative evaluation using NLP techniques over real-world examples of posts and responses of users.

We begin our analysis by gaining a global understanding of the problem. Specifically, we intend to know "what are the most frequently discussed topics when it comes to rape and sexual abuse." To answer this question, from every post in our database, we extract bi-grams and tri-grams from the responses of users and calculate their frequency across all posts. In order to compare these two subreddits, Table 4.1 and Table 4.2 report the bi-gram and tri-gram for both datasets. In Table 4.1 which specifically for /r/rape subreddit, unsurprisingly, we see that words such as *sexual assault*, *sexual abuse*, and *sexual violence* are the most frequently discussed words. That being said, we also observe some interesting words that pertain to entities such as *friend*, *mom, dad*, and *ex-boyfriend* which reveals the type of people who play an important role in a victim's situation and recovery. People also seem to frequently use words that convey emotions such as *feel safe*, *feel comfortable*, and *experiencing dealing outside*. In addition to this, words such as *help*, *talk*, and *please help* show that people acknowledge Reddit as a secure platform to seek help for stigmatized issues like rape.

As for /r/rapecounseling n-gram results listed in Table 4.2, we can also see some bi-gram shows the same exact results such as *sexual assault*, *need help* and *mental health*. This subreddit also mention important people in their discussion as if *best friend* and *family member*. Despite the similarity from the results in Table 4.1, there are very clear differences in tri-gram results in Table 4.2. It shows that the discussion in

Table 4.1: Most Frequent n-grams in /r/rape Subreddit Posts

Bigram	Trigram
sexual assault	rape sexual assault
months ago	local rape crisis
sexual abuse	taking time write
makes feel	perspective decision confident
ex boyfriend	person moment unfortunately
started dating	lot internal keeping
sexual violence	happened months ago
hard time	experiencing dealing outside
multiple times	deleted completed lot
feel safe	becoming person removed
panic attack	zine submission encrypted
little bit	zine experiences anonymously
mental health	wrote removed thank
feel guilty	vulnerable able word
feel comfortable	voice creating resource
weeks ago	vocal married abuser
sex life	violence publicly discussed
rape sexual	victims sexual violence
please help	told mom dads
kept saying	time taboo sex

Table 4.2: Most Frequent n-grams in /r/rapecounseling Subreddit Posts

Bigram	Trigram
feel like	accusing someone circle
sexual assault	used shun someone
first time	lashback accusing someone
best friend	suffer doubt victim
don't remember	rape crisis center
rape kit	don't really know
mental health	never told anyone
false accusation	don't know feel
want sex	think lots victims
need help	think people violate
sexual abuse	people violate deny
get away	public social media
think people	feel like deserve
someone else	victims suffer maybe
social media	police admit friend
victim blaming	victim blaming culture
sorry happened	wanted tell people
feel guilty	support action evidence
feel comfortable	assault prosecutor never
family member	selfish people compared

/r/rapecounseling subreddit focused on the action that should be taken by the victim such as *police admit friend* and *support action evidence*. The discussion scope also shows how responses among users indicate that social media is a medium for them to share. We found these tri-gram results for /r/rapecounseling highlight the importance of social media as seen in the phrases *public social media* and *rape crisis center*. This finding shows that the user also considers /r/rapecounseling subreddit as a crisis center for them to refer when it comes to rape. User response reveals the common reluctance to share problems with actual people outside of the Reddit community by mentioning *feel comfortable*, *accusing someone circle*, *never told anyone*, and *selfish people compared*.

In addition to that, we would like to learn how open are victims in expressing their issues to the outside world. We studied the responses of victims that exclusively talk about topics such as *share, talk* and *discuss* to find the following examples from both subreddits:

- "*I want as little people to know as possible, so people not constantly walking on eggshells around me.*"-/r/rape
- "*Although it was nice to talk to a third party about all of this, but I didn't feel like it was fixing the problems I'm having.*"-/r/rapecounseling

From the above responses, one can observe that victims have the chance to express their issues in Reddit. Moreover, Reddit also provides a secure and anonymous atmosphere, which the real-world fails to provide.

4.3 Analysis of Semantic Categories

Our second goal is to understand whether responses are always emotional and sympathetic. To answer this, we use the Linguistic Inquiry and Word Counts package

(LIWC) [5] , which is a comprehensive tool that leverages thousands of emotional and psychological word dictionaries to map the input words into 64 semantic categories (Park *et al.*, 2012; Coppersmith *et al.*, 2015c; Pennebaker *et al.*, 2015; Coppersmith *et al.*, 2015a). Table 4.3 and 4.4 show the top semantic categories returned by LIWC for both subreddits which paired with the words that pertain to the corresponding categories. To compare these subreddits, we extract the most frequent word from the same categories. Obviously, we see categories such as *sad* and *swear*, which signifies strong emotion and sympathy; however, we also observe some inconspicuous categories such as *money* and *religion.* From Table 4.3, we see that *money* is frequently associated with words such as *cost, wealthy* and *greedy*, while *religion* is associated with *hell, doom*, and *confession.* With that said, Table 4.4 shows a slightly different result for /r/rapecounseling subreddit. For the *money* category, we can see the associated words mentioned financial issue such as *payment, expense, affordable.*

Table 4.3: List of Words from Top Semantic Categories for /r/rape

Semantic Category	List of words
Sad	cry, suffer, miss, hurt, ruined, isolated
Home	roommate, neighbor, family, resident, homesick, bed
Swear	sh*t, c*nt, pisses, idiot, b*st*rd, b*llsh*t
Money	payment, expense, cheapest, affordable, income, poverty
Religion	spirit, doom, confession, god, priest, pray
See	hazy, stare, appearance, reveal, vivid, disappear

[5]http://liwc.wpengine.com/

Table 4.4: List of Words from Top Semantic Categories for /r/rapecounseling

Semantic Category	List of words
Sad	lose, alone, missing, resigned, tears, depression
Home	lease, rental, domestic, renovation, nanny, neighbor
Swear	freaky, retarded, moron, sh**head, f**kboy, screw
Money	payment, expense, cheapest, affordable, income, poverty
Religion	praying, sacrificing, afterlife, holy, blessing, immorality
See	disappear, selfie, revealing, depiction, stare, appearance

To get a deeper understanding of these issues, we used the words associated with money and religion (i.e., Table 4.3) as search queries to retrieve responses that pertain to these queries. When it comes to money-related issues, we found that responses were mostly about the cost associated with therapy and seeking professional help. For example, consider the following responses that we extract from both rape and rape counseling subreddit:

- "*Seeing anyone professionally isn't a financial possibility, I cannot take anti-depressants even if I had money for a prescriber and meds.*"-/r/rape
- "*I'll make sure to take your advice through and focus on finding a therapist which helps me alleviate these problems.*"-/r/rapecounseling

When it comes to religion, we found that victims usually deal with conflicting emotions due to confounding theories between religion and sex. The following posts illustrate this observation.

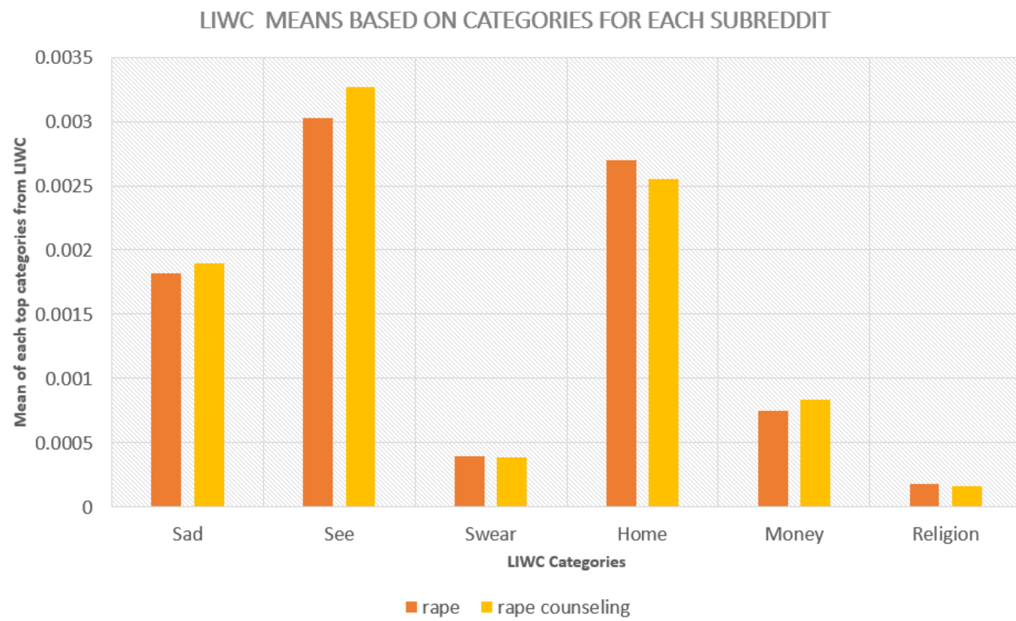


Figure 4.1: Emotion Attributes for Subreddit /r/rape and /r/rapecounseling.

- "*My girlfriend is abusing my religion and me, I feel like suicidal.*"-/r/rape
- "*I try to find God. I know religion isn't for everyone but my spirituality really does help me. I hope you are ok.*"-/r/rapecounseling

Furthermore, our third goal is to compare /r/rape and /r/rapecounseling subreddit. We select top semantic categories from Table 4.3 and 4.4. At first, we calculate the mean value for both subreddits. Then, we use a hypothesis test in order to compare two independent population proportion. Since /r/rape and /r/rapecounseling subreddits have the same scope of discussion, we want to see if these two datasets are different from each other. In order to differentiate /r/rape and /r/rapecounseling subreddits, we apply statistical t-test by setting the *alpha* value to 0.05. A significance value (p-value) of the difference is reported. The p-value is the probability of obtaining the observed difference between the samples if the null hypothesis were true. The null hypothesis will be rejected if the p-value is less than or equal to 0.05.

Figure 4.1 displays the mean scores of the six top semantic categories as extracted by LIWC from Table 4.3 and 4.4. Based on two-sample t-test, there were no notable differences in *religion* and *swear* categories across the two subreddits. That means, in the scope of *religion* and *swear* categories, a user from both subreddits expressed a similar level of these emotions. However, *sad*, *see*, *home* and *money* had different pattern. Users in rape counseling subreddit were more likely to express emotion in *sad*, *see* and *money* categories. Meanwhile, for rape subreddit, the *home* category shows the higher mean score. For each emotion attribute, we use t-test to check if the /r/rape subreddit distribution is not significantly different from those of /r/rapecounseling subreddit and the null hypothesis is rejected if p-value ≤ 0.05.

4.4 Extracting Useful Topics

Our fourth goal is to extract a broad range of topics that are relevant to the victim's post (or query). To achieve this, we train the LDA topic model to retrieve topic distributions related to rape. The input to LDA is a set of user posts (i.e., victims' question), and each post is a thread of comments by the respondents and the number of topics was set as 100 and the hyperparameters alpha and beta were set to 0.01 and 0.001 respectively (Blei *et al.*, 2003; Amir *et al.*, 2017; Manikonda and De Choudhury, 2017). Our goal is to retrieve topics that are relevant to the user's query of interest. To achieve this, the following steps were adopted. First, we manually selected some important queries from the actual postings of users. We select queries in both Table 4.5 and Table 4.7 were for representation purpose specific to our problem. For instance, Reddit data that we crawled contain several noisy words and some are too generic. If we automatically choose a seed word, generic words such as "rape" and "abuse" might come up. However, these are not useful in our scenario. On the other hand, words such as "emotional support" or "insurance" is more useful and hence, looking for LDA topics that contain such words enables us to provide useful and impactful examples to the readers. This is the main reason we designate certain important issues (or words) from user's posts (i.e., subreddits and their replies). Second, we used these query words as search tokens to filter out topics that had an exact match with the query in the top five ranked list of words. These topic clusters provided us with more relevant words that were related to the query words.

Table 4.5: Linguistic Topic Distribution using Latent Dirichlet Allocation (LDA) on /r/rape Subreddit

Query	Top words
"Therapy and insurance stuff giving me headache"	relationship, money, depression, health, benefit
"I'm having anxiety and trauma after that incident"	people, therapist, rape, support, assault
"Looking for support from sexual abuse"	symptom, PTSD, illness, mental, duty
"Should I take legal action for my justice"	case, court, police, guilty, trial
"I don't report because of lack of evidence"	victim, evidence, crime, domestic service

Table 4.5 and Table 4.7 show results for /r/rape and /r/rapecounseling subreddit respectively. The first column shows the actual user query and the underlined set of words indicate the keywords in these queries and the second column shows the top

Table 4.6: Sample Sentences for Top Words in Table 4.5

Sentences
“You have every right to interview as many therapists as you want before deciding on who’s right to help you. If you have insurance, that’s a place to start.” “You deserve a relationship in which the other person treasures you andseeks to validate and support you.”
“Some day, perhaps. Relatives and friends can be very forgiving, especially if they know that you’re still going through a hard time and need their support.” “What happened to you was a serious crime and a trauma and you may find that you need some help in moving past it.”
1) “You shouldn’t put your mental health at risk by teaching consent to a guy who’s already shown he’s not interested in learning about it.” “If you continue to feel bad about it, or have other symptoms like bad dreams or flashbacks (it can definitely be a side effect of medical assault.”
“Legally, the RCCs aren’t allowed to discriminate, but it’s true that they’re oriented more toward the needs of female victims.” “Don’t become what you see yourself as in your traumatized, guilty and altered mindset.”
“When someone commits a crime, if their victim doesn’t realize it, it’s still fully the fault of the criminal.” “I’d say it may not be worth re-traumatizing yourself to go forward with a trial where there’s no physical evidence to back it up.”

words that are relevant to these query terms. Furthermore, Table 4.6 and Table 4.8 are sentences that we extracted from the actual response from our dataset. For instance, in Table 4.5, when we searched for *anxiety* and *trauma*, words such as *therapist* and *assault* were produced by the topic model. Meanwhile in Table 4.7, *frustration* and *depression* top words been produced from the same query. Therefore, such relevant words were in-turn used as search tokens to query the model recursively. It is interesting to see that the *discussions of people are not confined to emotional support*. In this case, we observe that user response also touches on a specific topic such as insurance. This is quite surprising for us because previous studies by De Choudhury and De (2014); showed that people's discussions are limited to emotional support. On the contrary, our results reveal that insurance is also one of the main concerns when it comes to a health-related issue such as getting sick, medical treatment, therapy, etc. Similarly, legal advice is another interesting topic we discovered through our methodology. Note that some query for both subreddit were similar and some are different. This is because we only generate the top topic for each subreddit. Legal is one of the top discussion in /r/rape subreddit but it was not the top topic in /r/rapecounseling. Aside from that, we can see in Table 4.7 that user discuss more on trying to recover and finding the right resources for professional help in /r/rapecounseling subreddit. These results show that these two subreddits are not exactly similar in terms of topic distribution.

4.5 Recommending Relevant Solutions

Finally, we are ready to understand the ramifications of these diverse set of topics. To achieve this, we use topical words (see the second column of Table 4.5 and 4.7) to retrieve the actual responses of users that match with these topics for each subreddits. The results of this outcome are depicted in Table 4.6 and Table 4.8 respectively,

Table 4.7: Linguistic Topic Distribution using Latent Dirichlet Allocation (LDA) on /r/rapecounseling Subreddit

Query	Top words
"Therapy and insurance stuff giving me headache"	extent, finance, feeling, deserve, exhale
"I'm having anxiety and trauma after that incident"	abuser, cheer, life, frustration, depression
"Getting mental health treatment is tricky"	luck, supportive, justified, panic, understated
"Seriously looking for specialist/professional for my depression issue"	quick, problem, drama, specialist, flasback
"I need a solid plan on my recovery process"	delusion, course, adult, wishing, irritable

where one can observe the responses related to issues such as relationship, therapy, court, justice, specialist, treatment, and recovery. For example, when it comes to queries on therapy in Table 4.5, respondents suggest the victims explore many options before deciding on the right therapist in Table 4.6. Meanwhile, in Table 4.7, users express her/his own experience when getting help. More importantly, when it comes to topics on *legal advice, justice* and *evidence* there seem to be many confounding responses with discussions talking about the validity of physical evidence, discrimination between male and female sexes, and conflicting opinions on whether to proceed with the trial or not. The user also tries to get as much information from the discussion for the recovery process as we can see in Table 4.8 when respondent

Table 4.8: Sample Sentences for Top Words in Table 4.7

Sentences
"Those fights were about anything; from my finances to me just not coming home to see them enough (I came home almost every 2 weeks)..." "I think you're right about the exaggerating thing as well; so many times I walk out of therapy still feeling like I've understated myself..."
"I had an episode of psychotic depression when I first left my abuser- but it was misdiagnosed as a full blown psychotic episode..." "It's so easy to just bottle up and take out your frustrations on someone who doesn't deserve it..."
"If you're in the UK, hopefully, you should get referred to the secondary treatment team..." "...my closest friends have been incredibly supportive..."
"Hopefully if you tell them you've been having flashbacks, panic attacks, irritable, nightmares- these for over 2 years- that they should recognise it as PTSD and they SHOULD refer you to a specialist." "I recommend a quick google of self-care techniques- breathing has been good for me..."
"I occasionally hear voices to do with those delusions..." "I was raped repeatedly over the course of a year and no one knew about it"

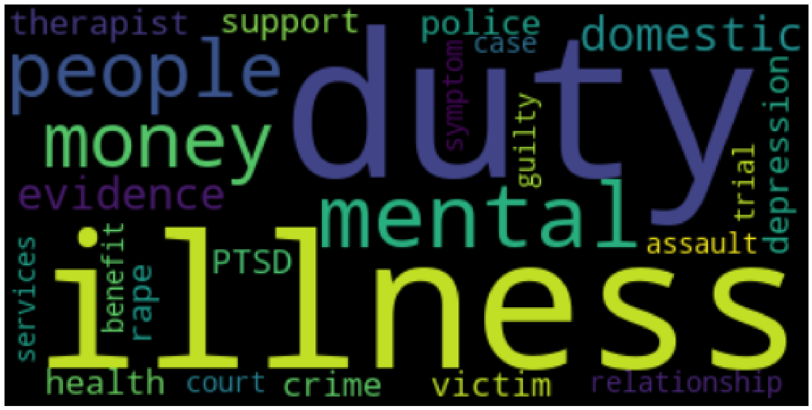


Figure 4.2: Word Cloud for Subreddit /r/rape Based on the Top Word Generated by LDA.

suggests several solutions. This clearly shows that user responses are diverse and not limited to emotional support.

In addition to that, we also visualize the top words generated by the topic modeling using a word cloud. This is to show if there exist outlier from these two datasets that we use. Figure 4.2 and figure 4.3 showed the visualization based on the top words that generated by LDA for both subreddit. In-depth, major discussion topic in r/rape as we can see in figure 4.2 are mostly related to legal action. On the other hand, figure 4.3 that represent r/rapecounseling shows that user trying to get more information on the healing process. These two subreddits involve a different kind of discussion topics but there is a possibility of an existing outlier such as topic that exists in both

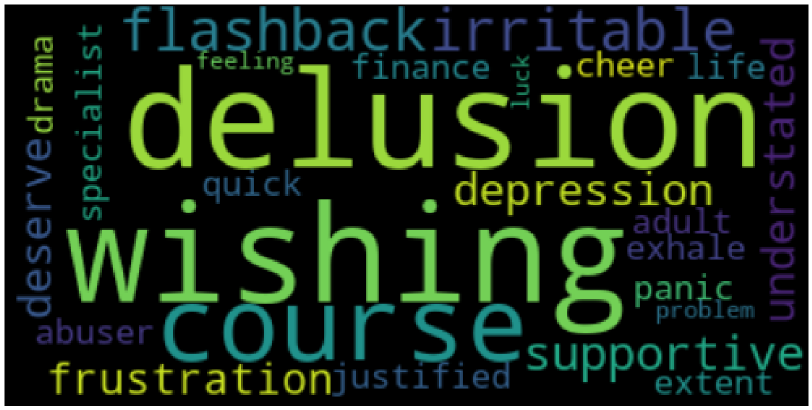


Figure 4.3: Word Cloud for Subreddit /r/rapecounseling Based on the Top Word Generated by LDA.

subreddits. We can see that words such as *"depression"* and *"support"* can be found in both subreddit showing that these topics are very common in both communities. Both dataset also shows that user uses the online discussion to get the support that they need especially emotional support. This happened because of the main scope of these subreddits is about a similar concern which is "rape".

4.6 Sentiment Analysis

Sentiment analysis is an extraordinary capacity which has profound implications to understand human behavior (Reagan *et al.*, 2017). It is an automated process of understanding an opinion about a given subject from written or spoken the lan-

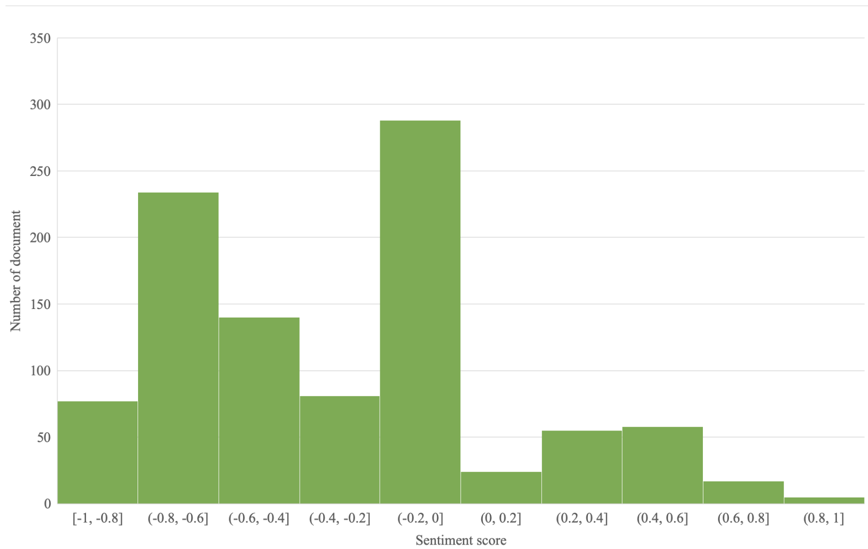


Figure 4.4: The Sentiment Score Distribution on Subreddit /r/rape

guage. The diversity and multiple dimension of human behavior make it very difficult to find out the exact sentiment from a text. It requires proper sentiment words to be extracted from the set of text. In this experiment, we want to study the sentiment of these two community by comparing the sentiment score for /r/rape and /r/rapecounseling subreddit. We use Vader(Valence Aware Dictionary and Sentiment Reasoner) sentiment analysis tool to get the score for data that we have (Gilbert, 2014). VADER is a lexicon and rule-based sentiment analysis tool that is specifically accustomed to sentiments expressed in social media. The sentiment lexicon generated by VADER is sensitive both the polarity and the intensity of sentiments expressed in social media. There are different types of polarities in texts: positive evaluation, negative evaluation, and neutral evaluation.

Figure 4.4 and Figure 4.5 shows the sentiment score distribution for /r/rape and /r/rapecounseling subreddits respectively. VADER generates four scores which are positive sentiment, negative sentiment, neutral and compound. The compound score is computed by summing the valence scores of each word in the lexicon, adjusted

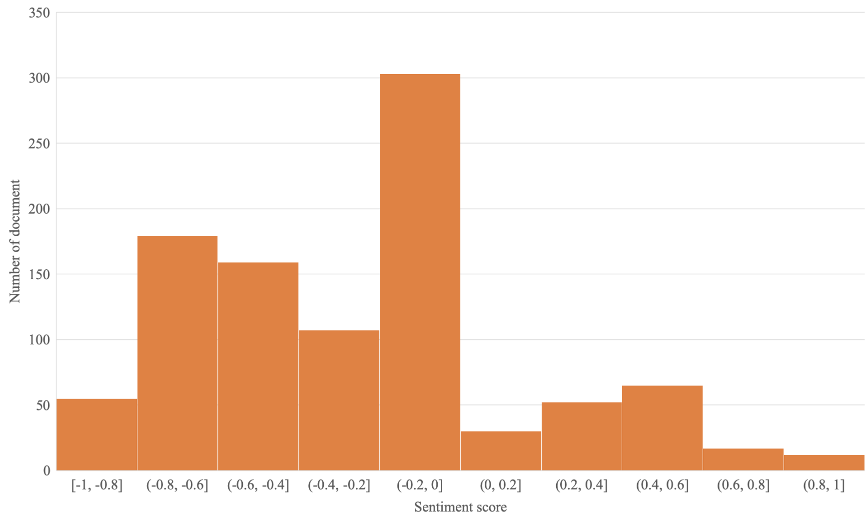


Figure 4.5: The Sentiment Score Distribution on Subreddit /r/rapecounseling

according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). We selected the compound score in order to do this comparison. We found that the result in Figure 4.4 which contain the sentiment score for /r/rape subreddit was higher in negative score compare to Figure 4.5. For example, we can see from the second interval comparing Figure 4.4 and 4.5, there is a reduction in the negative score. This finding explains that how /r/rapecounseling subreddit provide less negativity in the discussion and responses are more likely on solution and support towards users. This is because /r/rapecounseling include professional in their discussion thread which contributes to less negative sentiment across the discussion.

In addition, we run another experiment on sentiment classification for both subreddits. Sentiment classification can be conducted at the document, sentence, or phrase (part of sentence) level (Abbasi *et al.*, 2008). For our experiment, we conducted on a sentence level by taking posts by the user. Then, all sentences were automatically labeled with -1(negative), 0(neutral), and 1(positive). Many studies

Table 4.9: The Classification of Posts from Subreddit /r/rape based on Sentiment Score.

Classifier	Accuracy	F1
Naive Bayes	85.19%	61.38%
Logistic Regression	84.14%	60.41%
Random Forest	83.76%	45.48%

Table 4.10: The Classification of Posts from Subreddit /r/rapecounseling based on Sentiment Score.

Classifier	Accuracy	F1
Naive Bayes	84.38%	65.90%
Logistic Regression	85.54%	67.15%
Random Forest	84.62%	48.45%

have used machine learning algorithms with support vector machines (SVM) and Naive Bayes (NB) being the most commonly used. For this experiment, we choose NB, Logistic Regression (LR), and Random Forest (RF) classifier. Data were divided into 80% training and 20% testing. In order to evaluate machine learning models that we chose, we implement cross-validation resampling procedure. The average accuracy and F1 score over 10-fold cross-validation were compared for all three classification models.

Table 4.9 and Table 4.10 shows the classification results for both subreddits based on the sentiment score. The sentiment classification is responsible for categorizing an opinion into either positive or negative opinion from all posts from both subreddits. We were able to achieve the highest accuracy and F1 score with Naive Bayes for r/rape, 85.19% average accuracy and 61.38%. On the other hand, we achieved

85.54% average accuracy and 67.15% average F1 scores for r/rapecounseling reported by logistic regression classifier. Our experiment result on sentiment analysis does not clearly show major differences for both subreddits but we can still point out a slight difference. This is because opinions or sentiments are easy to understand by human beings, but it is not that easy for a computer to achieve the same level of understanding especially when the scope of the discussion is very close in context.

After all, results finally answer our question number three by showing that /r/rape was higher in negative sentiment compare to /r/rapecounseling. From this observation, we can see that the nature of discussion in those communities is slightly different. Discussion in previous sections also confirms certain characteristics of these subreddits that clearly present some contrast.

Chapter 5

USER ACTIVITIES ON MENTAL HEALTH COMMUNITIES

Mental health is and continues to be a prominent plague for the civilized world. It is estimated that one in four American citizens suffer from a diagnosable mental disorder in any given year (Nadeem, 2016). Health information seeking and information sharing on social media has relatively attracted researchers in this field to explore further in studying human behavior. It is still a very new and fragile field that can still be explored in many dimensions. On the other hand, the social media platform provides a really rich ecosystem to study social support and human behavior.

In this chapter, we extend the work on various online mental health community. We crawl textual data from 52 subreddits in order to understand the nature of user behavior specifically on the online mental health community. Health information seeking and information sharing on social media has relatively attracted researchers in this field to explore further in studying human behavior. It is still a very new and fragile field that can still be explored in many dimensions. Mental health has become a public concern nowadays. People have started to think about the importance of mental health problems and their effects on our society. This is not a minor issue; on the contrary, it is a very serious issue that can contribute to mental well-being. Researchers in the psychology field have studied this topic for decades. With the increasing use of social media data, this specific problem also attracts the attention of many computer scientists. Users actively share and communicate with online communities and researchers have found that it is a smart idea to leverage social media data to study this problem in order to help online communities and authorities at the same time.

This can contribute to an immense change in overcoming this issue. In this section, we address the following questions:

(1) How reliable online discussion can be in terms of helping users with the mental health problem?

(2) What is the sentiment distribution across online mental health community?

(3) Does all mental health community only have negative sentiment?

(4) What is the topic variation across community?

We crawled 52 mental health-related subreddits which specified in Table 5.1. We classified all subreddits into 5 categories by following to the previous work by Sharma and De Choudhury (2018). Those subreddits were classified into 5 types of online mental health communities (OMHC) using clustering and human labeling. They basically used a k-means clustering algorithm to perform the initial clustering on the n-grams (n = 3) of the post that is shared on all collected subreddits. Then, using two annotators that were familiar with mental health subreddit, they refined the machine-labeled clusters and all subreddits were assigned to suitable descriptive labels known as a community as presented in Table 5.1. We employed several text preprocessing techniques in our data collection to reduce noisy and unreliable data. The crawl of the subreddits used in this chapter spanned between May 20, 2018, and July 20, 2018. We elaborate on some descriptive statistics of our crawled data. Our data contained 37,606 posts with at least one comment on each post.

5.1 Sentiment Analysis

This section will discuss sentiment analysis on the OMHC subreddit. The diversity and multiple dimensions of human behavior make it very difficult to find out the exact

Table 5.1: Five Reddit OMHC Categories and Their Associated Subreddits Sharma and De Choudhury (2018)

Community Category	Community	#Posts
Trauma & Abuse (C1)	r/abuse, r/adultsurvivors, r/afterthesilence, r/Anger, r/bullying, r/CPTSD, r/domesticviolence, r/emotionalabuse, r/ptsd, r/rapecounseling, r/StopSelfHarm, r/survivorsofabuse, r/SurvivorsUnited, r/traumatoolbox	10399
Psychosis & Anxiety (C2)	r/Anxiety, r/BipolarReddit, r/BipolarSOs, r/BPD, r/dpdr, r/psychoticreddit, r/MaladaptiveDreaming, r/Psychosis, r/PanicParty, r/schizophrenia, r/socialanxiety	8661
Compulsive Disorders (C3)	r/calmhands, r/CompulsiveSkinPicking, r/OCD, r/Trichsters	3005
Coping & Therapy (C4)	r/7CupsofTea, r/Existential_crisis, r/getting_over_it, r/GriefSupport, r/helpmecope, r/hardshipmates, r/HereToHelp, r/itgetsbetter, r/LostALovedOne, r/offmychest, r/MMFB, r/Miscarriage, r/reasonstolive, r/SuicideBereavement, r/therapy	9284
Mood Disorders (C5)	r/depression, r/depressed, r/ForeverAlone, r/GFD, r/lonely, r/mentalhealth, r/Radical_Mental_Health, r/SuicideWatch	6731

sentiment from a text. It requires proper sentiment words to be extracted from the set of text. We use Vader sentiment analysis tool to get the score for data that we have (Gilbert, 2014). VADER is a lexicon and rule-based sentiment analysis tool that is specifically accustomed to sentiments expressed in social media. The sentiment lexicon generated by VADER is sensitive to both the polarity and the intensity of sentiments expressed in social media. There are different types of polarities in texts: positive evaluation, negative evaluation, and neutral evaluation. VADER generates four scores which are positive sentiment, negative sentiment, neutral and compound. The compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive).

We begin by analyzing the header of each subreddit that we have from our dataset. To understand the sentiment of the mental health community, we start by analyz-

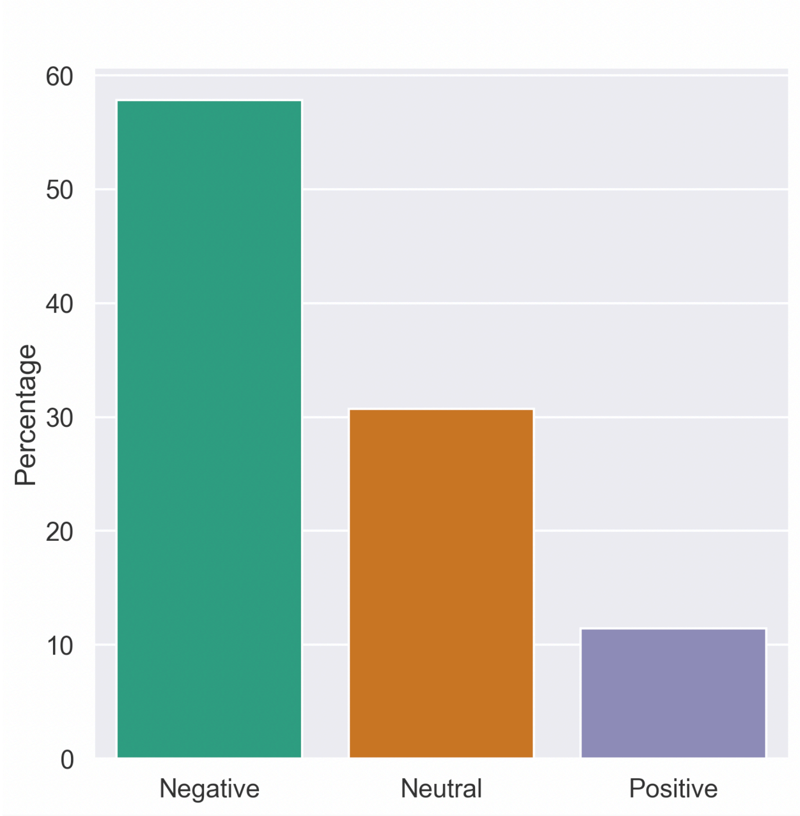


Figure 5.1: The Sentiment Score on “abuse” Subreddit from C1

ing the discussion headlines of each subreddit. The sentiment score for each headline from 56 mental health-related subreddits were measured (Sharma and De Choudhury, 2018). Then, we identify which subreddit was supposed to be a support subreddit discussion and which subreddit should be just a discussion on a specific topic in general. In Table 5.2, we present the sample headlines that we extract from our dataset. Refer to figure 5.4, headlines sentiment score for this subreddit has a higher negative sentiment. This shows that the subreddit was unreliable if an upcoming user only

Table 5.2: Headlines of a Few Mental Health Oriented Reddit Posts.

Does anyone else have this issue? I often feel very lonely, as I mostly by myself
20 year old sister is abusive towards our 50 year old father
I need tips on where to go for help.
After 5 years of anxiety depression and daily panic attacks I've decided enough is enough!
Anybody else up? I'm losing my mind in isolation.
Anyone else? I am obsessed with making sure that I have a regular level of morality and concerned I'm Evil.
Appropriate resentment or misplaced feelings?
Am I an abuser and if so what should I do?

looks at the subreddit headlines to find support from the discussion. From our understanding, this subreddit should provide support discussion to users but the sentiment score was higher on negative sentiment. It shows how user discussion might lead to negative sentiment. Referring to other subreddits in figure 5.1 and figure 5.5 also reported higher negative sentiment. On the other hand, figure 5.2 and 5.3 have higher neutral sentiment. These subreddits that supposed to have higher negative sentiment scores produce higher on the neutral sentiment which is unexpected. Through our experiment, we found that subreddit headlines were not reliable to study the sentiment across the mental health community. This means that positive subreddit might contain negative sentiment discussion and vise versa. So, to study the sentiment distribution across the mental health communities, using the headlines alone was not enough. The headlines can indicate a positive or negative sentiment but the

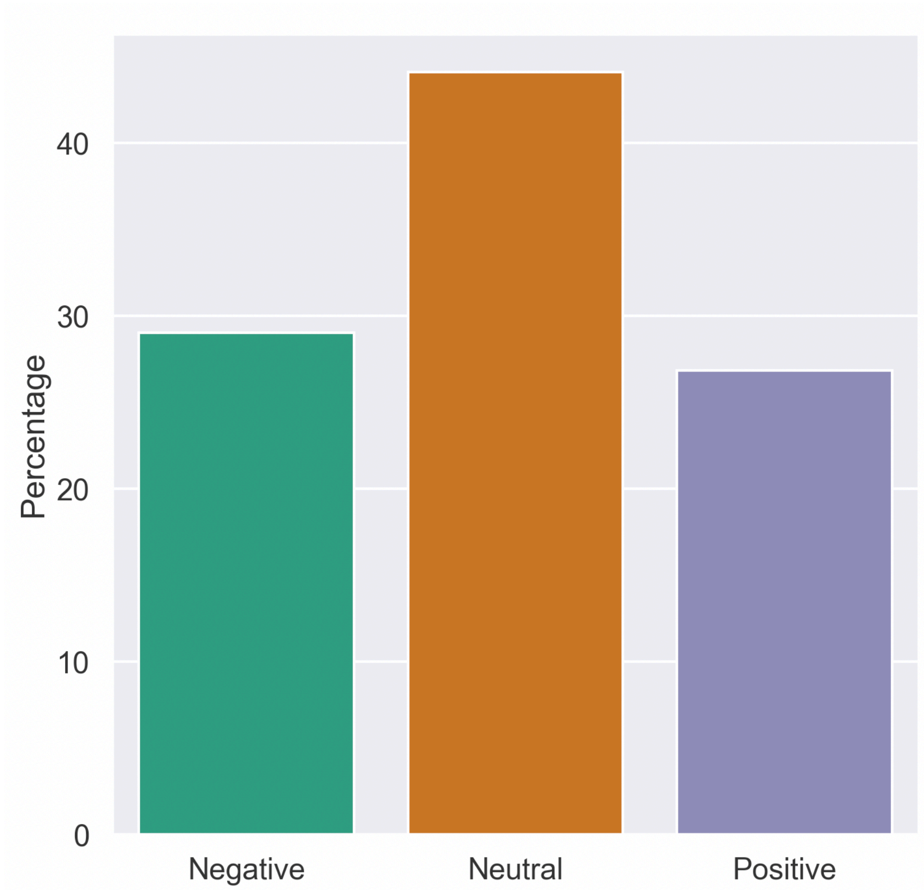


Figure 5.2: The Sentiment Score on "social anxiety" Subreddit from C2

discussion content will have a way different distribution. To answer our questions, we collect the whole discussion content of the selected subreddits starting from the headlines together with all of the comments associated with it.

Refer to Figure 5.6 for details sentiment analysis results for each mental health community. We observe that Figure 5.6 (a) which represents C1 is the only result with higher negative sentiment compare to positive sentiment. Other than that, all other community shows that positive sentiment is higher than the negative sentiment which is surprising. Based on these findings, the discussion content in C2, C3, C4, and C5 yields higher positive sentiment because these communities tend to discuss support and encouragement among each other. From the sentiment analysis result,

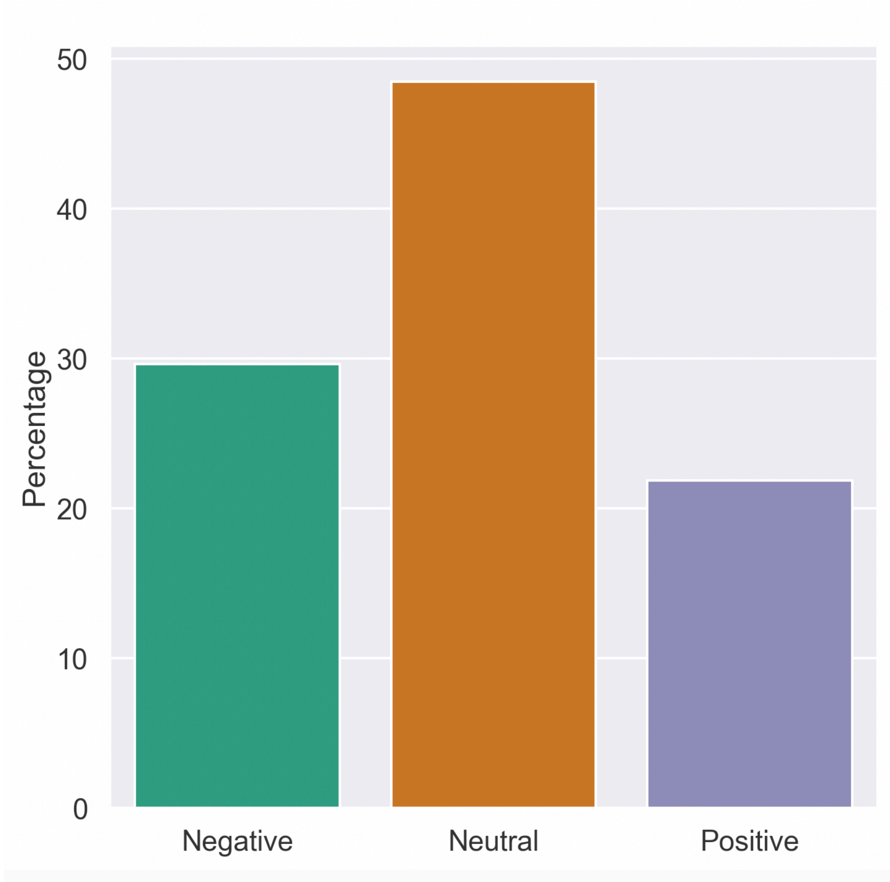


Figure 5.3: The Sentiment Score on "OCD" Subreddit from C3

we can see that the distribution of sentiment among online community sparse between "very negative" and "very positive". There are very few distributions in the middle of the graph. This is due to the topic discussed in the community. People tend to condemn people or support people. That explains the sentiment distribution.

5.2 Topic Distribution across Communities

To further study the mental health community, we run a topic modeling on our dataset on each community. To do that, we use Latent Dirichlet Allocation (LDA), a generative probabilistic model of a corpus that works by drawing distribution topic for each word in the document (Blei *et al.*, 2003). We empirically decided to extract

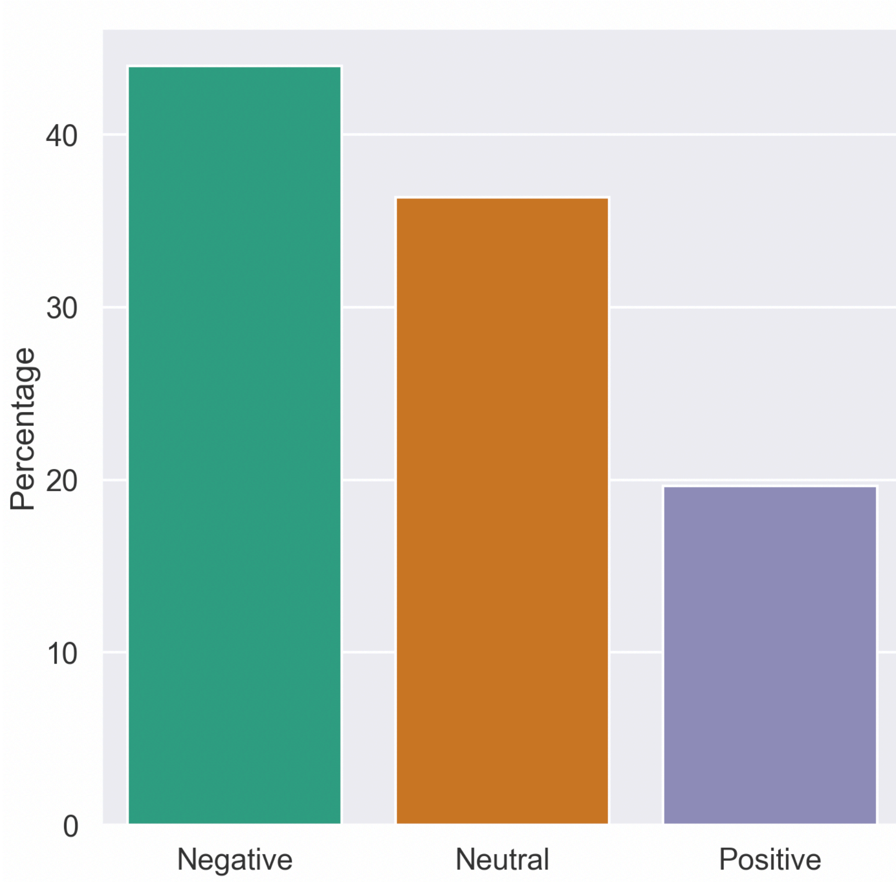


Figure 5.4: The Sentiment Score on "grief support" Subreddit from C4

5 topics from each community and their corresponding vocabulary is shown in Table 5.3. These topics reveal that different topic's distribution across communities. On the other hand, we also visualize the top words generated by the topic modeling using a word cloud for each community. This is to show if there exists an outlier and the scope of the discussion from these communities. Figure 5.7 through 5.11 showed the visualization based on the top words that generated by LDA for all communities.

Community 1 (C1) contains topics that are majorly discussed on the assault and abuse as we can see in the top words for Topic 1 in Table 5.3. We can see words such as *rape, abuse, and assault*. Users in this community also use that medium to get help and information. We can see the topic distribution that contains top

Table 5.3: Topic Distribution Across Mental Health Community.

Community 1	
Topic 1	sexual, remember, abuse, rape, memories, sorry, assault
Topic 2	study, research, link, survey, community, anonymous, help, government
Topic 3	people, trauma, experience,therapy, life, shame, symptom, emotional
Topic 4	help, live, leave, police, job, money, time, social, services
Topic 5	mental, doctor, medication, psychiatrist, insurance, treatment, depression, anxiety, diagnosed
Community 2	
Topic 1	people, person, life, love, sense,look, reality, time, feel, thinking
Topic 2	aphantasia, keen, agora, monologue, clenching, speculative, depleted, testimonies, malfunctioning
Topic 3	lithium, blood, anxiety, alcohol, stress, caffeine, life, supplement, dose
Topic 4	daydreaming, maladaptive, world, fantasy, character,music, reality, time
Topic 5	anxiety, panic, doctor, medication, feel, help, sleep,attack, dose, depression
Community 3	
Topic 1	obsession, compulsion, ocd, tried, lord, company, months, quiet, doubt, job, false, anxiety
Topic 2	hair, pulling, trich, head, scalp, bald, feel, eyelashes
Topic 3	feel, people, anxiety, time , life, bad, help, day
Topic 4	therapy, brain, people, experience, intrusive, mental, life, talk
Topic 5	reassurance, counselor, browsing, compulsion, argue, solely, seeking, fear, religion
Community 4	
Topic 1	life, feel, time. person, yourself, help, mind, help, hope, world, suicide, believe
Topic 2	women, helpful, wrong, rational, emotion, negative, yourself, sex, reason, bad, pretty, rape
Topic 3	baby, miscarriage, pregnancy, pain, sorry, doctor, bleeding, loss, feel, blood
Topic 4	people, marriage, post, media, social, care, husband, understand
Topic 5	illness, mental, creative, schizophrenia, narrative, obese, hobby, disease, happy, wallowing, struggle, cringy, delusion
Community 5	
Topic 1	exhaustion, allergic, construction, challenge, operate, compose, reflect, endorphin, machinery, destruct
Topic 2	message, thrawn, remindmebot, subject, compose tedious, empire, validation, rebel
Topic 3	humiliation, dream, guess, people, stuff, bullies, listen, rotten, school, fantasy
Topic 4	depression, help, anxiety, feel, mental, job, therapist, health, doctor, medication, money
Topic 5	life, family, parent, mom, dad, live, school, suicide, die, money, home

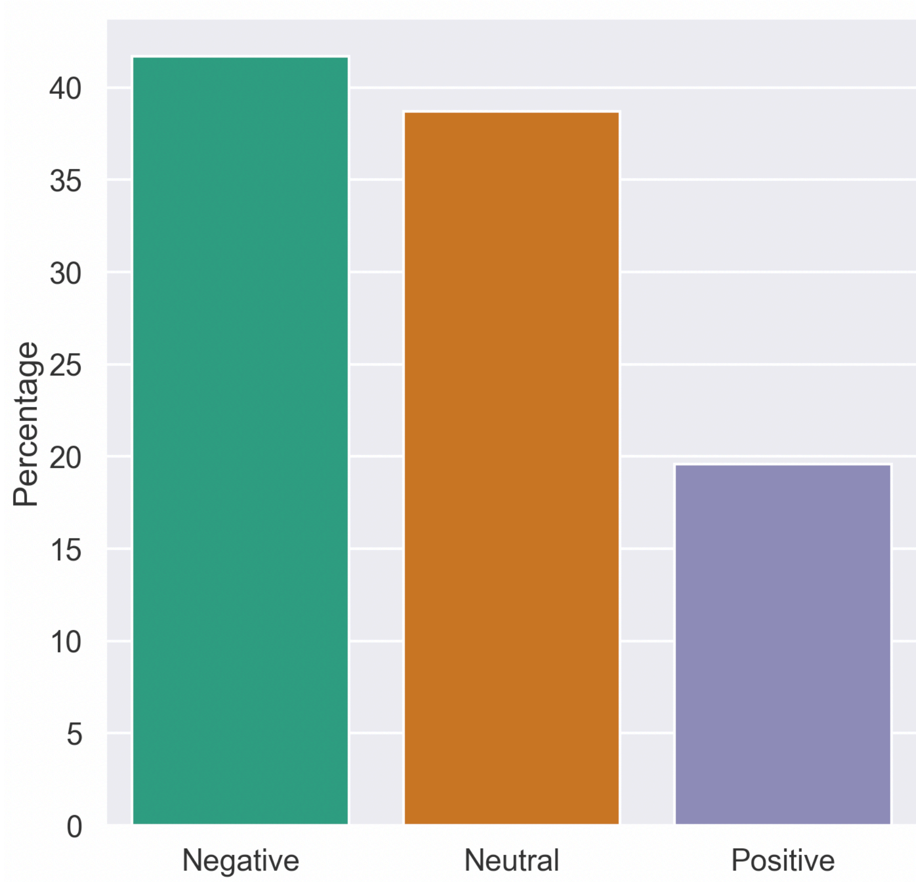
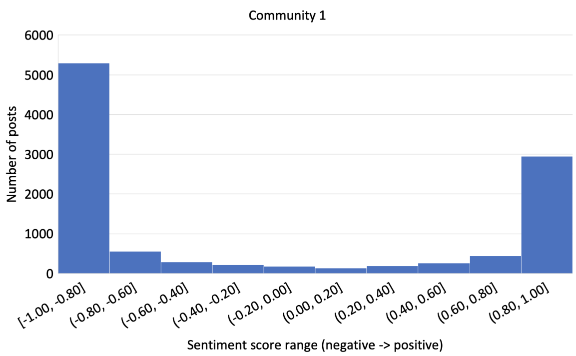


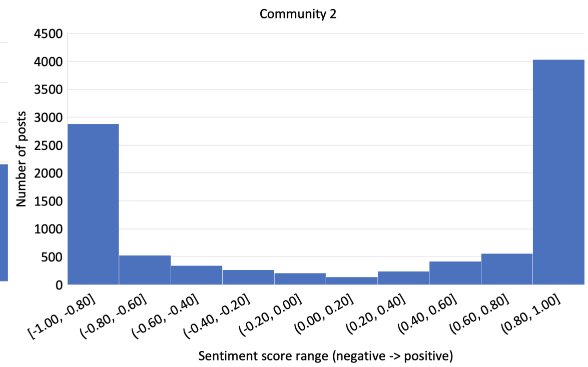
Figure 5.5: The Sentiment Score on "depression" Subreddit from C5

words like *help, therapy, treatment, government.* The topic distribution also shows the discussion among users in this community talked about money-related issues that we can found words like *insurance, money.* For community 1, the user more focused on discussing their symptoms and seeking information through from Reddit.

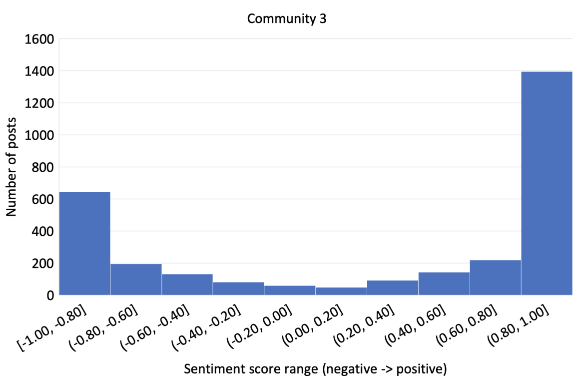
Moving on to community 2 (C2) topics extracted for this community shows that some groups of users discuss a topic related to mental images. Words such as *aphantasia, monologue, daydreaming* were observed from the topic distribution of this community as listed in topic 2. Users in this community also focused on getting treatment and discussing drugs. We can see in the list of top words for topic 3 which contained words such as *lithium, alcohol, supplement, dosage.* This shows that the user in this



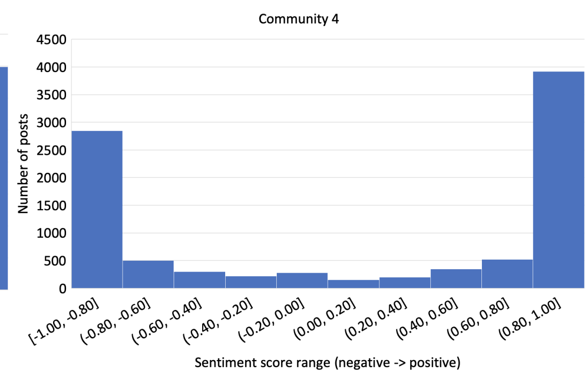
(a) The Sentiment Score for C1



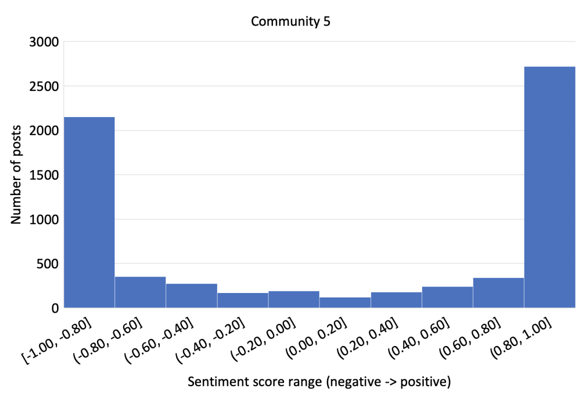
(b) The Sentiment Score for C2



(c) The Sentiment Score for C3



(d) The Sentiment Score for C4



(e) The Sentiment Score for C5

Figure 5.6: Sentiment Score for Each OMHC Community.

Figure 5.7: Word Cloud for Community 1 Based on the Top Word Generated by LDA.

community come to Reddit to share and get specific information. The topic distribution on this community also shows that users were trying to seek help based on the topic 5 top words *doctor, medication, help*.

On the other hand, community 3 which represents the compulsive disorder shows a variation from mental health action. The compulsive disorder community contains subreddits such as r/calmhands, r/CompulsiveSkinPicking, r/OCD, r/Trichsters. This group of discussions mainly about a syndrome to the disorder. We can observe from words like *obsession, compulsion, anxiety, hair, pulling, trich, scalp* as the top words in topics 1 and 2 which indicate the activity of users trying to explain their symptoms. Furthermore, this community also focused on the treatments as we can see from the top words in topic 4 and 5 that involve words such as *therapy, reassurance, counselor, talk*. In conclusion, this community talks more on the symptom of the disorder and treatment information.

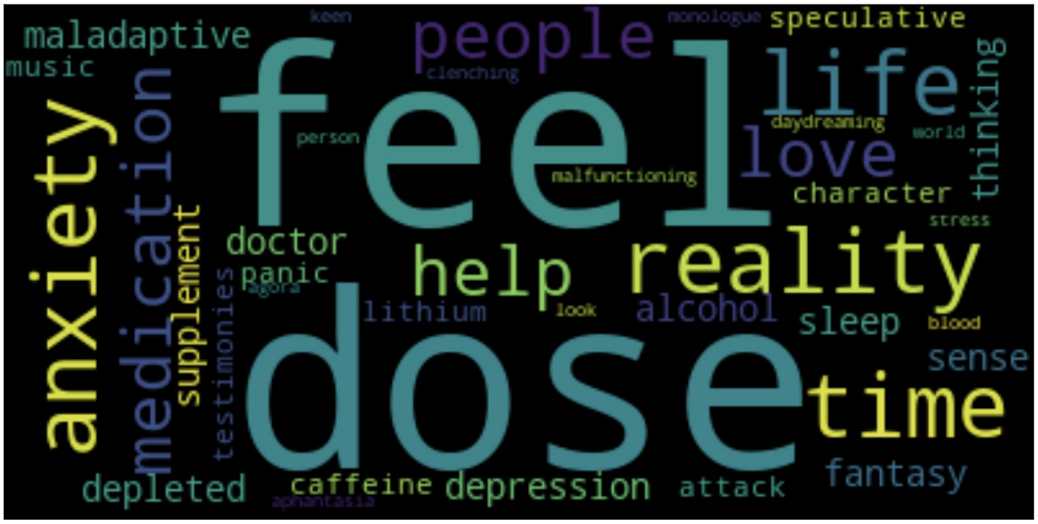


Figure 5.8: Word Cloud for Community 2 Based on the Top Word Generated by LDA.

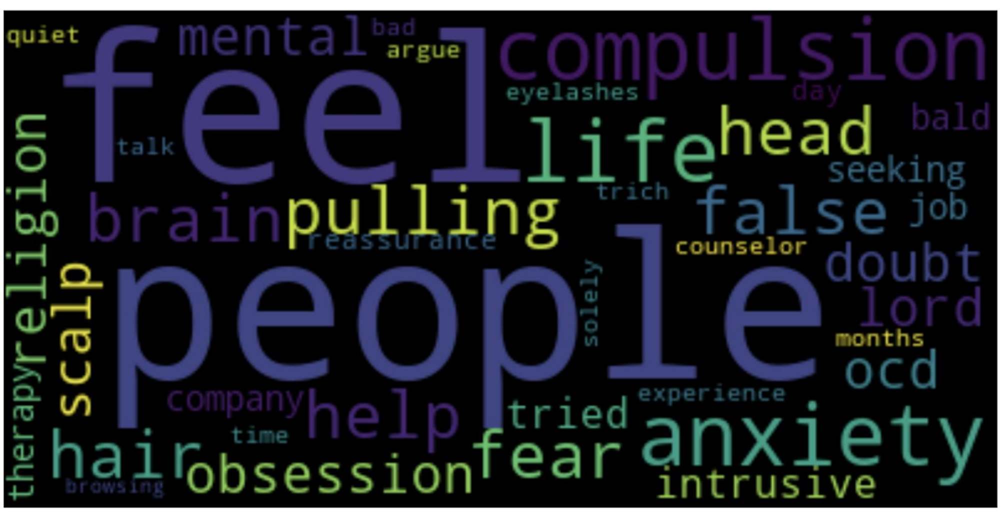


Figure 5.9: Word Cloud for Community 3 Based on the Top Word Generated by LDA.

Figure 5.10: Word Cloud for Community 4 Based on the Top Word Generated by LDA.

Besides, community 4 which represents a coping and therapy subreddits were more likely focusing on a topic that involves women's health. We can see words like *miscarriage, pregnancy, husband*. We also discover that this community also discussed the mental problem as we can see top words such as *delusion, schizophrenia, mental*. This community is less likely to discuss treatment because we did not have any observation on that. Lastly, topic distribution for community 5 shows that this community tends to discuss more on an intense feeling that we can observe words such as *depression, suicide*. People in this community also discuss lifestyle-related topics that we can see words like *life, money, job*. We can conclude that this community has a broad topic distribution.

The last community is community 5 which stands for mood disorder that include eight subreddits such as r/depression, r/depressed, r/ForeverAlone, r/GFD, r/lonely, r/mentalhealth,r/RadicalMentalHealth, r/SuicideWatch. Based on the topic distribution extracted, we found that this community talks about symptoms as we observe

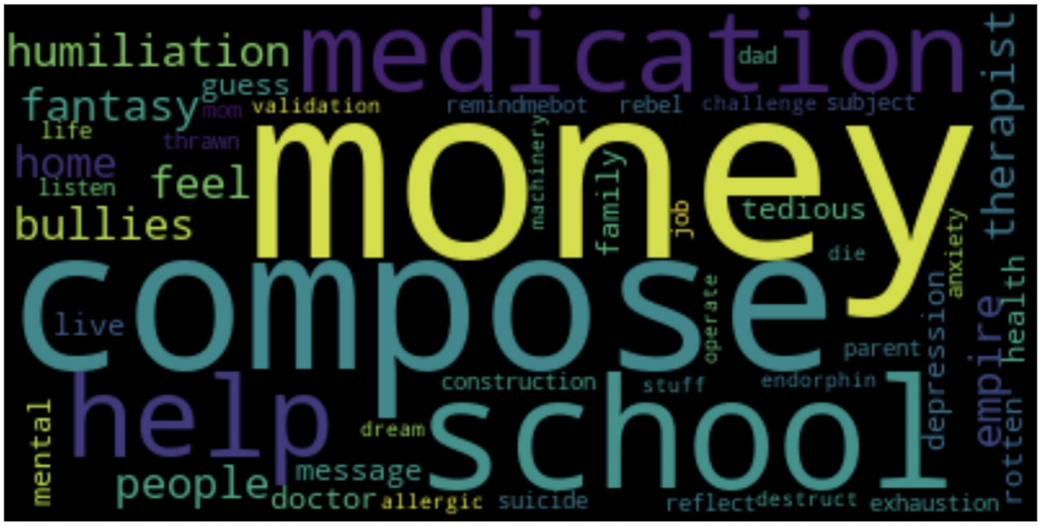


Figure 5.11: Word Cloud for Community 5 Based on the Top Word Generated by LDA.

top words like *allergic, exhaustion, dream, depression, anxiety, mental*. We also observe financial related topics and treatment were discussed in this community based on words like *money, job, therapist, doctor, endorphin, medication*. On the other hand, this community also reported a discussion that focusing on their family relationships such as *home, family, people, dad*. We can conclude that this community was mainly focused on the symptoms, treatment and financial discussion scope.

The similarities across all communities are each community will have a topic distribution on social support, feelings, and treatment which is not surprising. This supports the fact of the user does use an online forum such as Reddit as a medium for them to get support and information especially when it comes to mental health or other related stigmatized issues.

5.3 Correlation of Sentiment and Topic Distribution Results

In this section, we will elaborate on how sentiment results and topic distribution results are correlated with each other. The sentiment distribution for community 1 in Figure 5.6(a) shows that this community is higher on negative sentiment. This community is mainly focusing on trauma and abuse which it is expected to have higher in negative sentiment. At the same time, the topic that we extracted for this community showing that users in this community use the online forum medium to get help and information. They also discussed their symptoms which can contribute to the higher negative sentiment. This shows that results for sentiment and topic distribution for community 1 do correlate to each other.

On the other hand, the rest of the community which is communities 2, 3, 4 and 5 have a higher positive sentiment. These communities also reported that the communities did have a quite high negative sentiment but not as high as positive. For C2, the topic distribution majorly focusing on seeking help and getting medical information through an online forum. Meanwhile, the topic distribution for C3 is all about symptoms of their disorder and treatment information. For C4, this community's topic distribution is about intense feelings and even a lifestyle. These 3 communities show that users discuss a broad topic. This contributes to the lower negative sentiment and higher positive sentiment. For the last community (C5), the positive sentiment is higher but the negative sentiment distribution still a lit bit high compare to the other three communities. So, if we look at the topic distribution for C5, it shows that this community was mainly focusing on symptoms, treatments, and financial issues. This topic can contribute to a quite high negative sentiment as reported by Vader.

Based on our observation on both sentiment and topic distribution for all communities, results do correlate and confirms our analysis results on these communities are valid.

5.4 LIWC Emotional Attributes Across Communities.

To further understand the emotional attributes across the community, we use the Linguistic Inquiry and Word Counts package (LIWC) [1] , which is a comprehensive tool that leverages thousands of emotional and psychological word dictionaries to map the input words into 64 semantic categories (Park *et al.*, 2012; Coppersmith *et al.*, 2015c; Pennebaker *et al.*, 2015). We selected 6 top categories from a total of 64 semantic categories extracted by LIWC based on our previous work. We chose these six categories because we want to study the popularity of these emotion attributes among users. Kamarudin *et al.* (2018) also mentioned that these six categories are among the top ones when it comes to stigmatized issues. Figure 5.12, 5.13, 5.14, 5.15 and 5.16 show the LIWC distribution score for "sad" category across all 5 communities. Community 1 in Figure 5.12, community 2 in figure 5.13, and community 3 in 5.15 have similar score at 0.00 to 0.02 with the highest post frequency per each community. For community 3 in figure 5.14, a sligh lower score compare to the other three with the range of score in between 0.00 to 0.01. On the other hand, community 5 in figure 5.16 has the highest score range that falls at 0.00 and 0.03 with the post frequency above 6000 posts. From this score, we can see various distribution across the community. This is because each community carries a very different scope of discussion that may lead users to be less or more emotional.

Furthermore, Figure 5.17, 5.18, 5.19, 5.20 and 5.21 show the LIWC distribution score for "money" category across all 5 communities. Previous work has shown that

[1] http://liwc.wpengine.com/

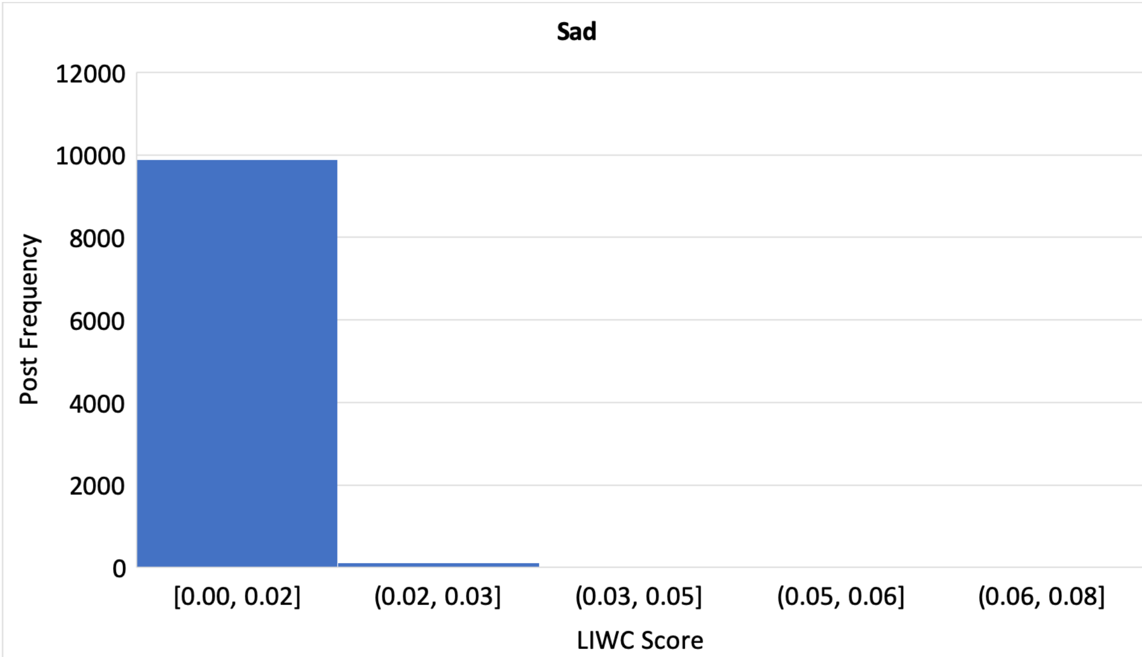


Figure 5.12: The “sad” Category LIWC Score for C1

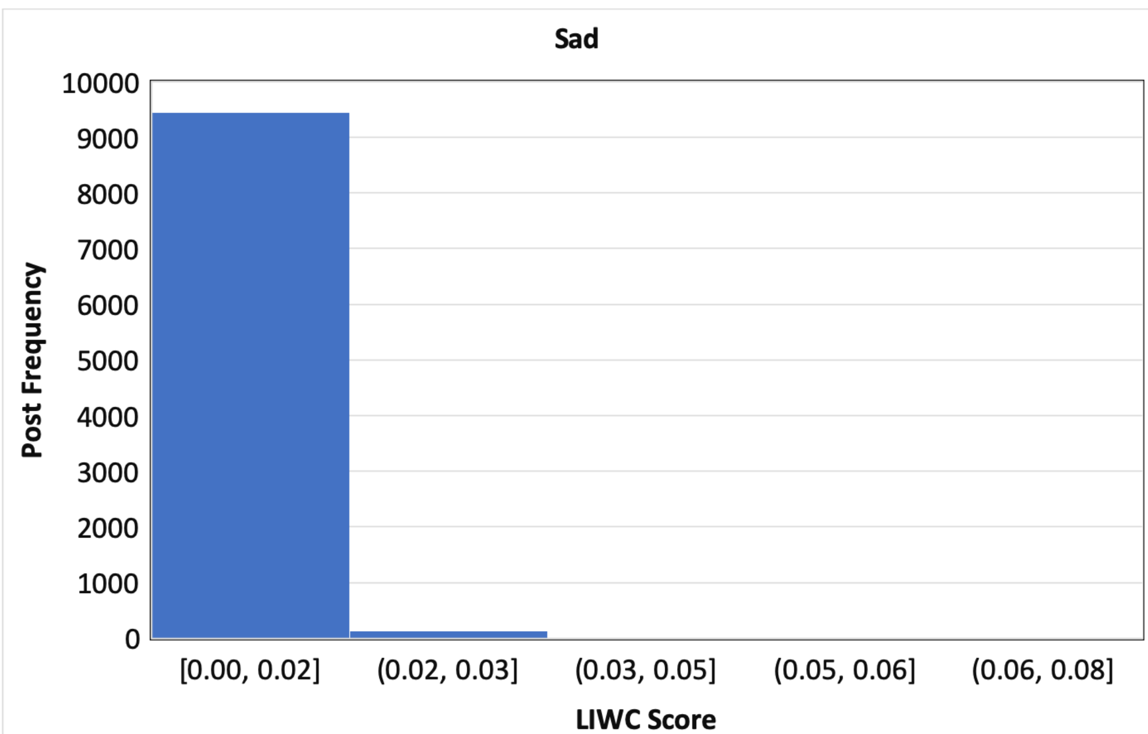


Figure 5.13: The “sad” Category LIWC Score for C2

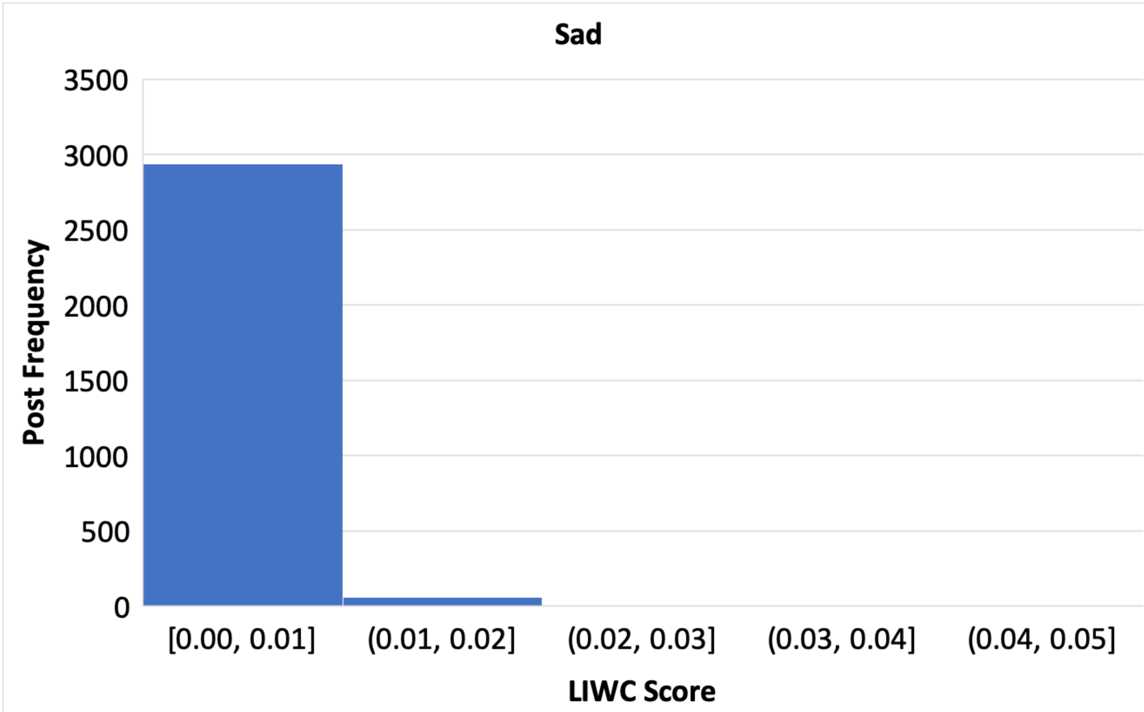


Figure 5.14: The “sad” Category LIWC Score for C3

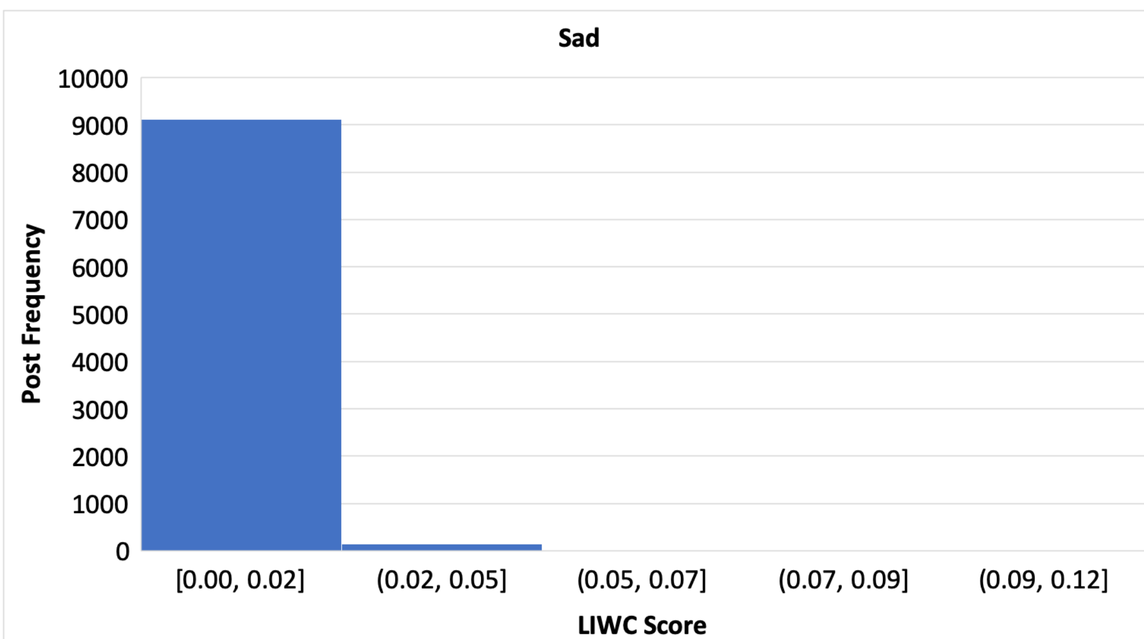


Figure 5.15: The “sad” Category LIWC Score for C4

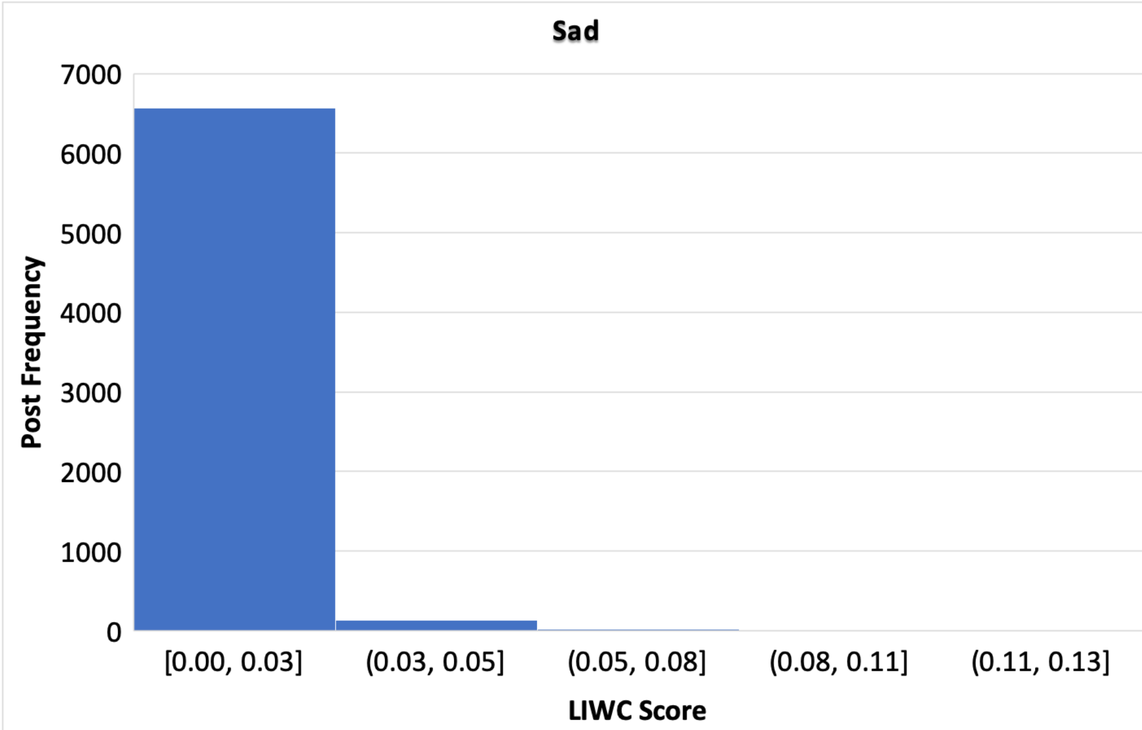


Figure 5.16: The "sad" Category LIWC Score for C5

this category is really useful because it shows that the scope of discussion among users is not limited to emotional support. It also shows that users are discussing various topic including insurance and a financial issue to get treated. We also reported the similar findings in our topic distribution that we discussed earlier in this chapter. For this experiment, we can see that C1, C2, and C3 share a similar range of scores for their majority posts with a range of 0.00 to 0.01. Meanwhile, C4 and C5 have the highest score range at 0.00 and 0.02 for their majority posts. This shows that communities 4 and 5 are leading with the highest score in terms of "money" category. It shows that people in this community discussed more on these issues.

In addition, figure 5.22, 5.23, 5.24, 5.25 and 5.26 shows the result for "swear" category for all communities. This community 1,2,3 and 5 have a similar range of LIWC score which is between 0.00 to 0.01. While only community 4 have a slightly

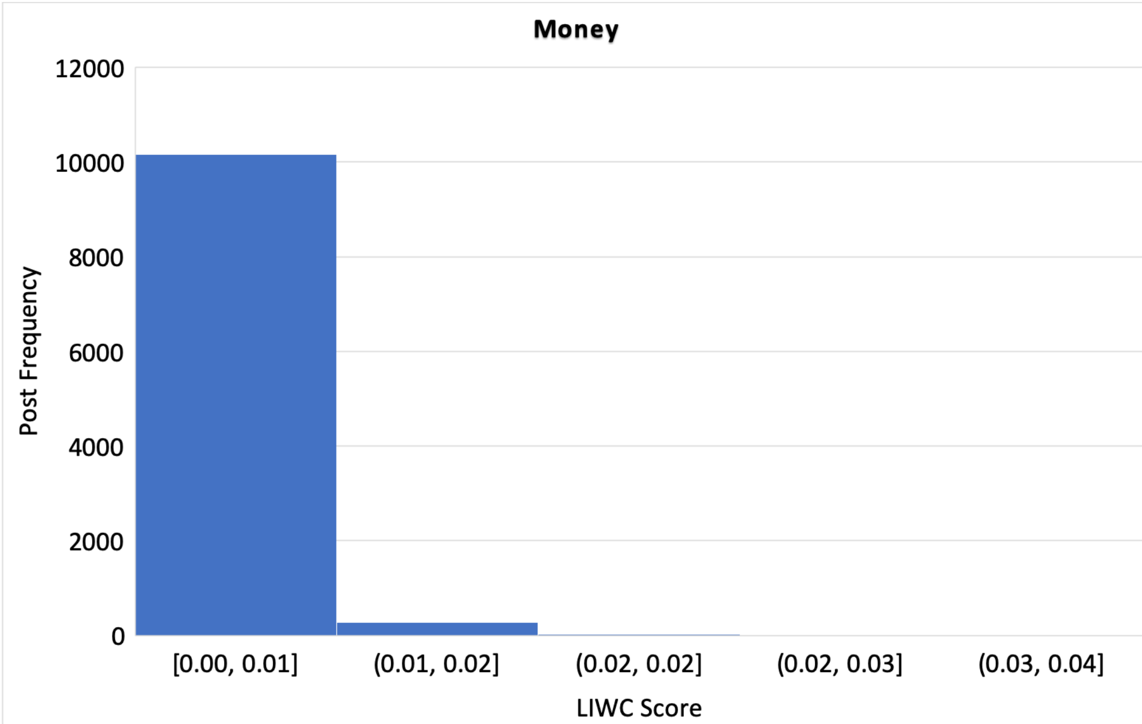


Figure 5.17: The "money" Category LIWC Score for C1

higher score for this particular category which ranges between 0.00 to 0.02. This result shows that the majority of the communities have a similar distribution on words that fall in the "swear" category as extracted by LIWC. This also shows that the use of words in this community is closely related to each other.

Other top categories that we selected based on our previous work is "religion". This specific LIWC category will select words that related to it and give a score for each post as to how other category works. Our results as depicted in figure 5.27, 5.28, 5.29, 5.30, and 5.32 shows that C1, C2, C3, and C5 have similar score range. On the other hand, C4 has a slightly higher score. These results are similar if we compare with the "swear" category. It shows that in terms of "swear" and "religion" categories, these 4 communities have a similar distribution of words.

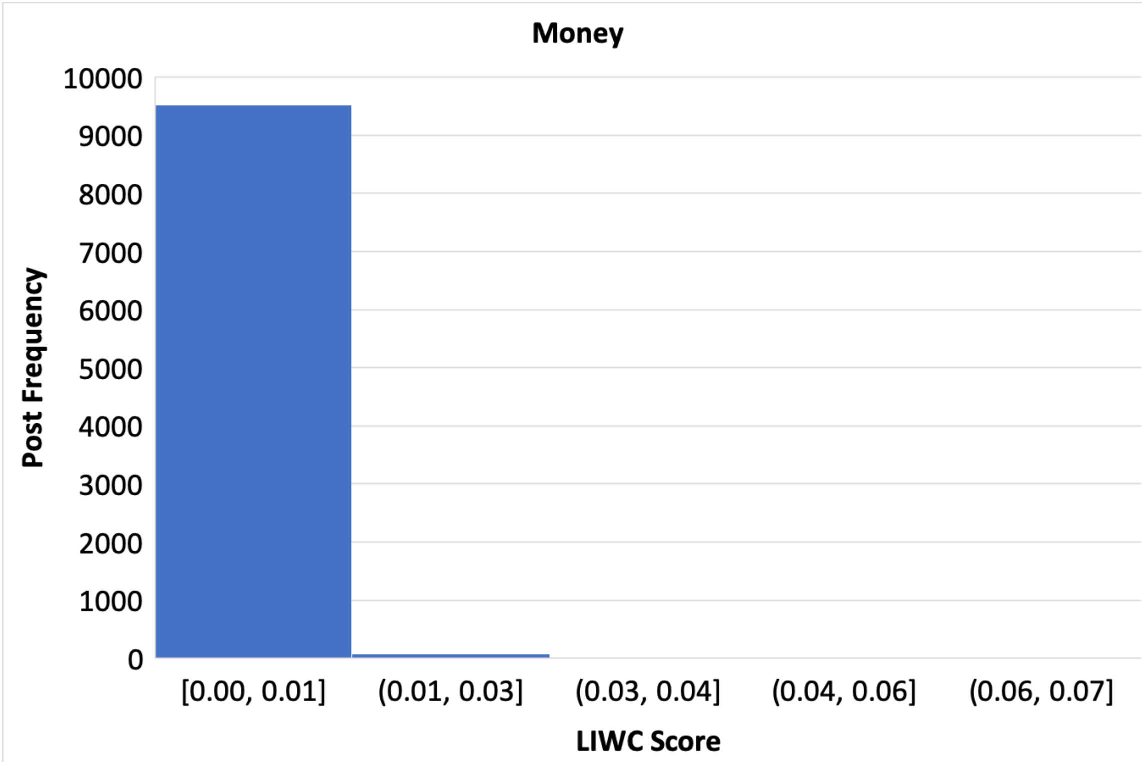


Figure 5.18: The “money” Category LIWC Score for C2

Moving on to another category known as “see” which represents one of the emotional categories by LIWC. This particular category will give a score to words that related to the emotional effect based on personality. For example, words such as *appearance, selfie, reveal* will specifically fall into this category. Since this category also one of the top categories, we extracted the LIWC score for each community. Figure 5.34, 5.35, 5.36, 5.37, and 5.38 display results all communities. As we can see from all these figures, C1, C4, and C5 have a similar range of scores between 0.00 to 0.02 with the highest post frequency. On the other hand, C2 and C3 share the same range with the highest score 0.03 for the majority of posts.

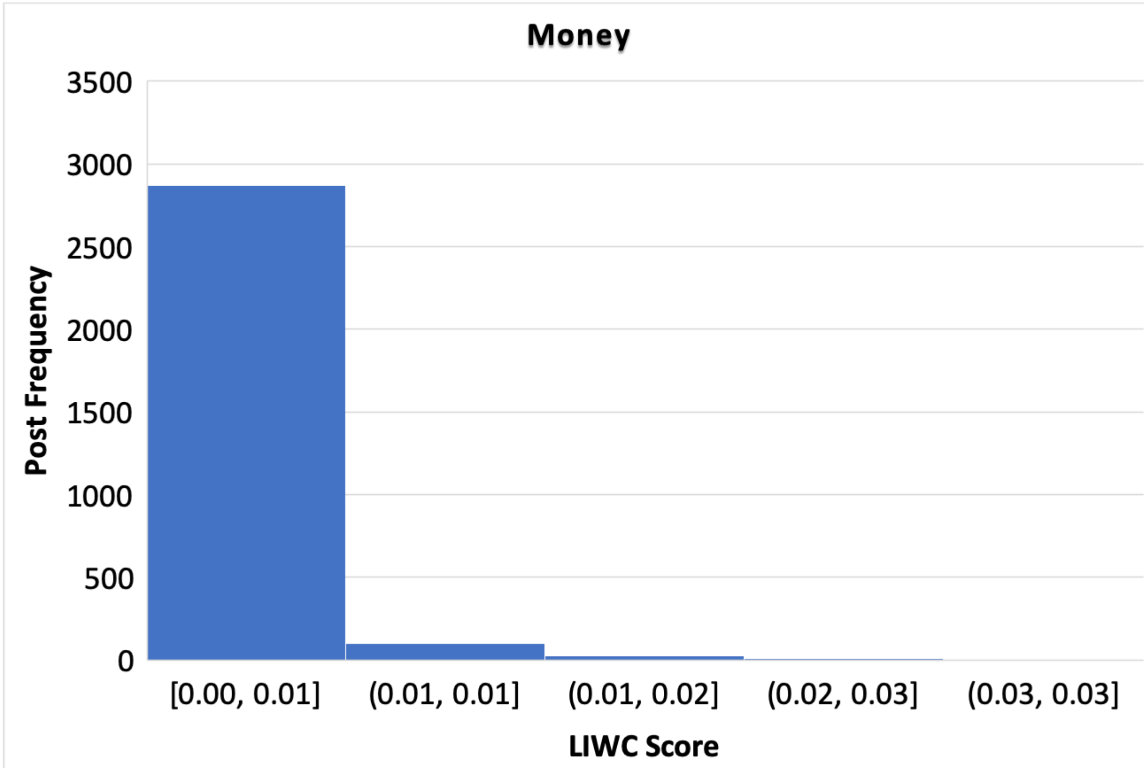


Figure 5.19: The “money” Category LIWC Score for C3

Another top category that we study from LIWC for this experiment is “home”. This category includes sample words like *roomate, domestic, rental, family*. This category will give a score to the community by counting the number of words mentioned in this category. We extracted the score for each community and reported the score in figure 5.39, 5.40, 5.41, 5.42, and 5.43. From our results, C1, C2, C4, and C5 share the same range of scores between 0.00 to 0.02 with their majority of posts fall in this range. Meanwhile, C3 is the only category that has a lower score range which is between 0.00 to 0.01. This shows that all 4 communities mentioned more words on this category compare to C4.

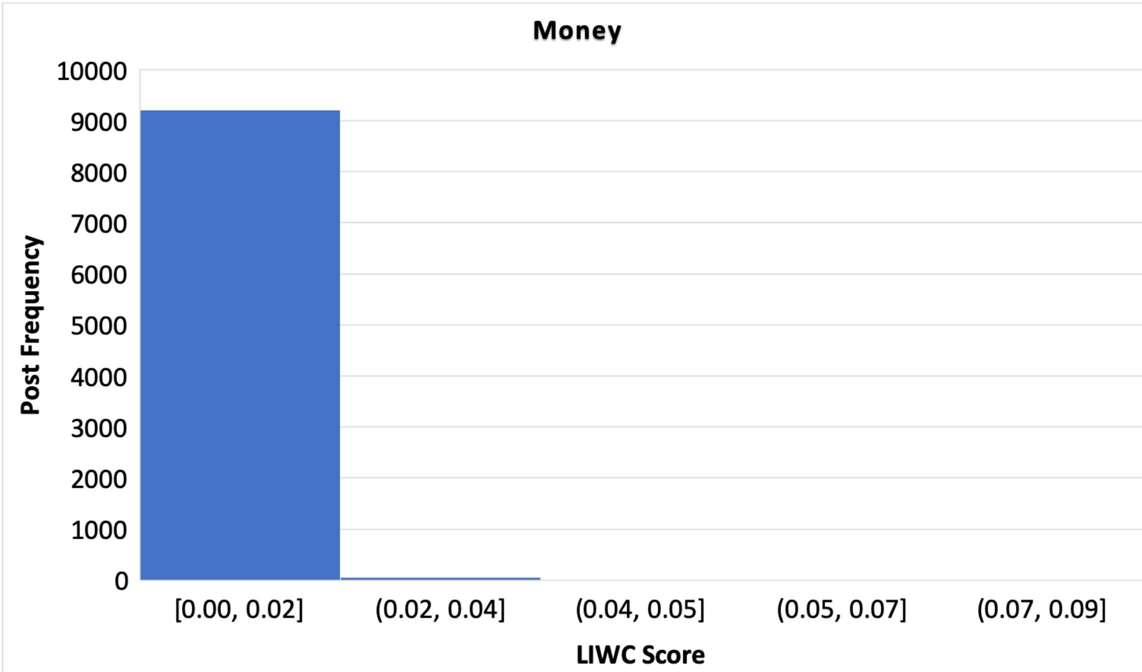


Figure 5.20: The “money” Category LIWC Score for C4

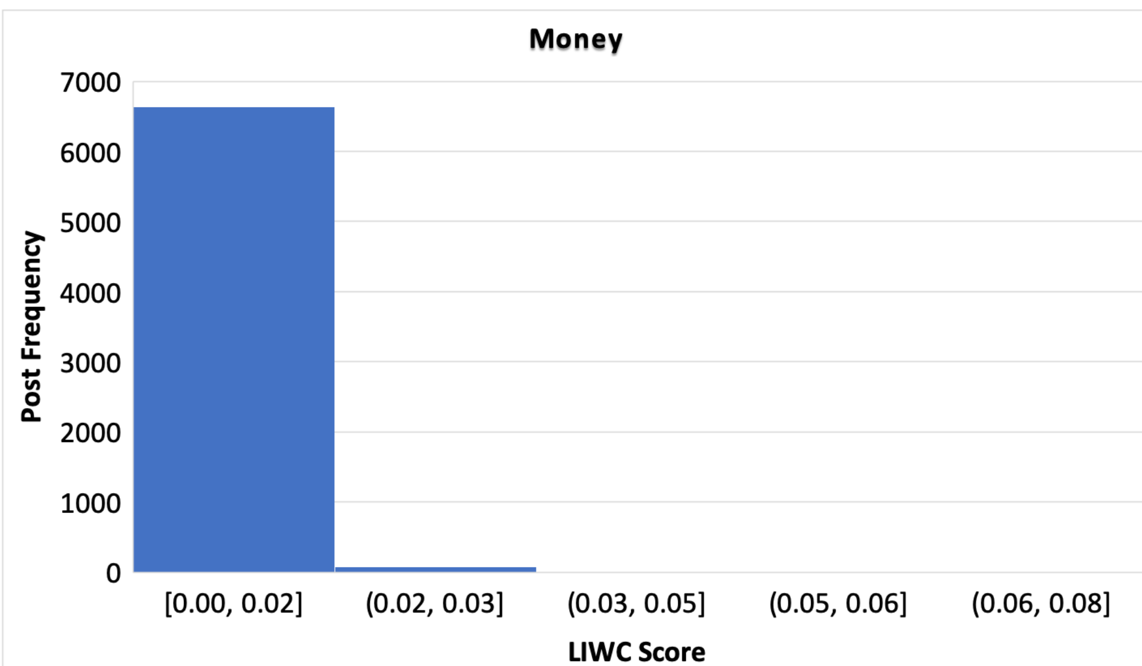


Figure 5.21: The “money” Category LIWC Score for C5

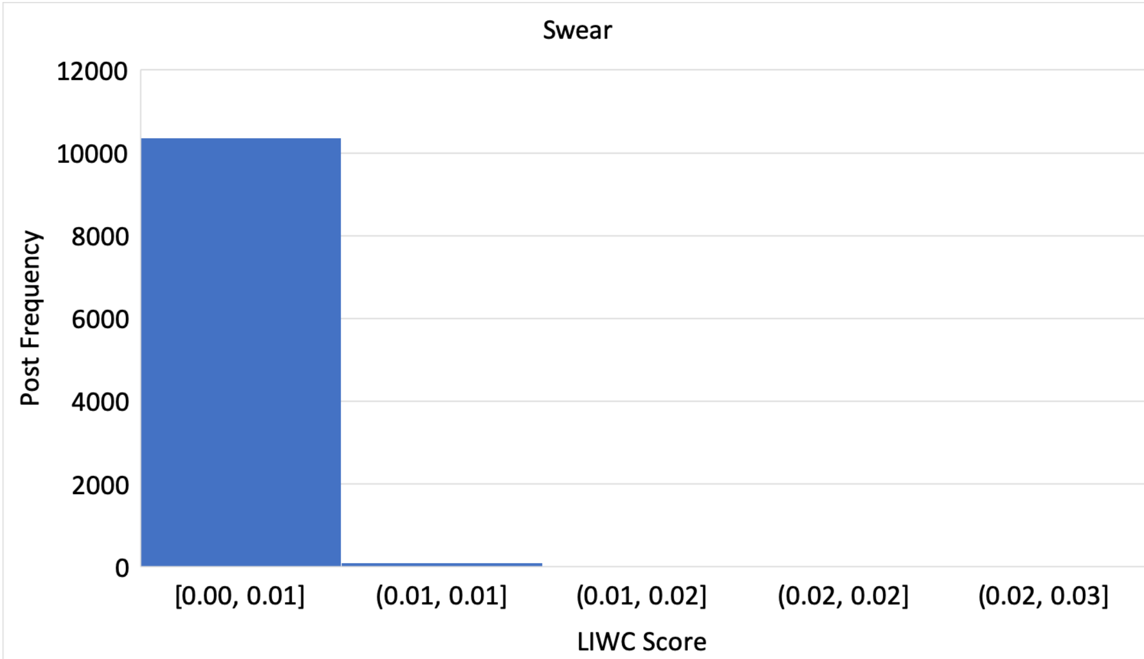


Figure 5.22: The “swear” Category LIWC Score for C1

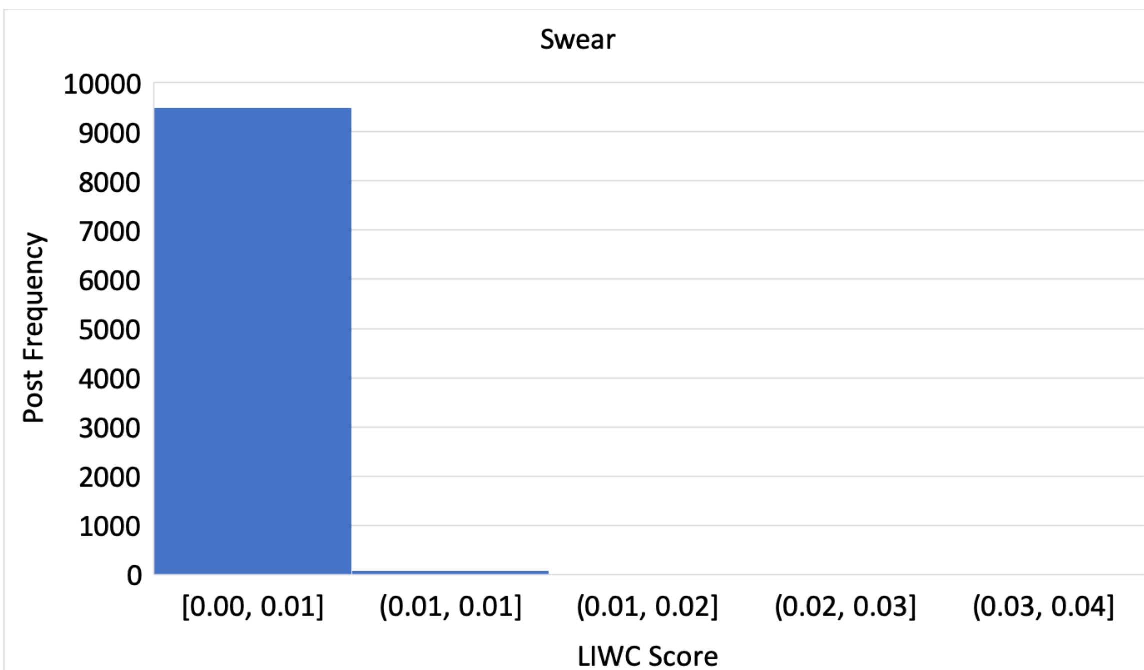


Figure 5.23: The “swear” Category LIWC Score for C2

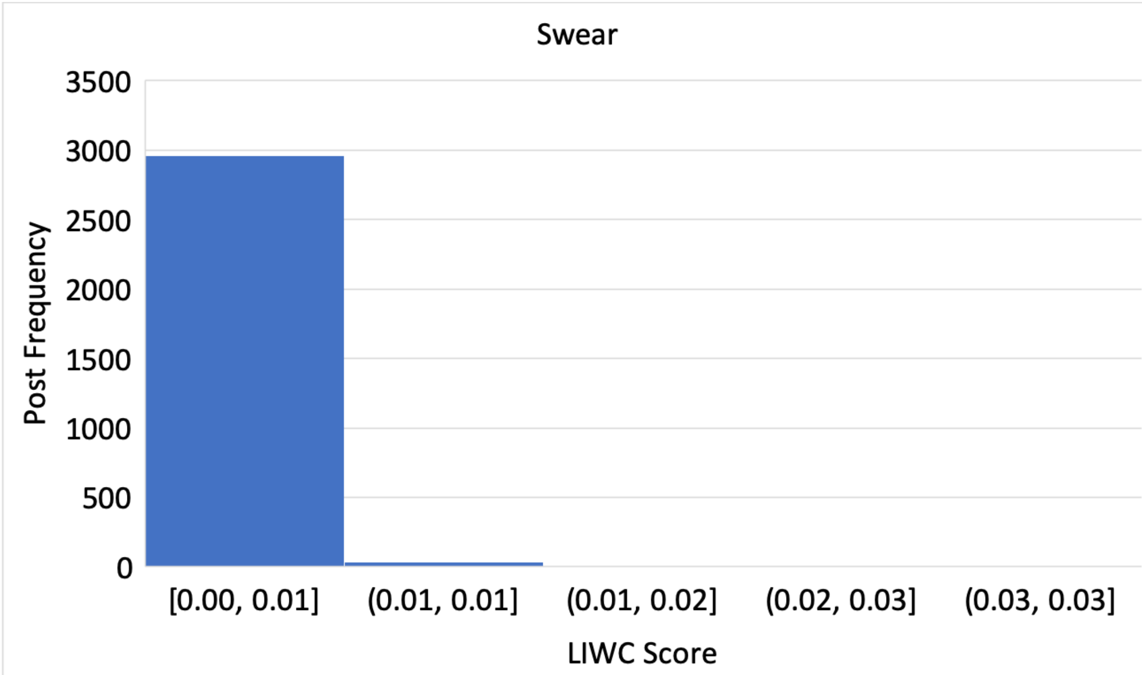


Figure 5.24: The “swear” Category LIWC Score for C3

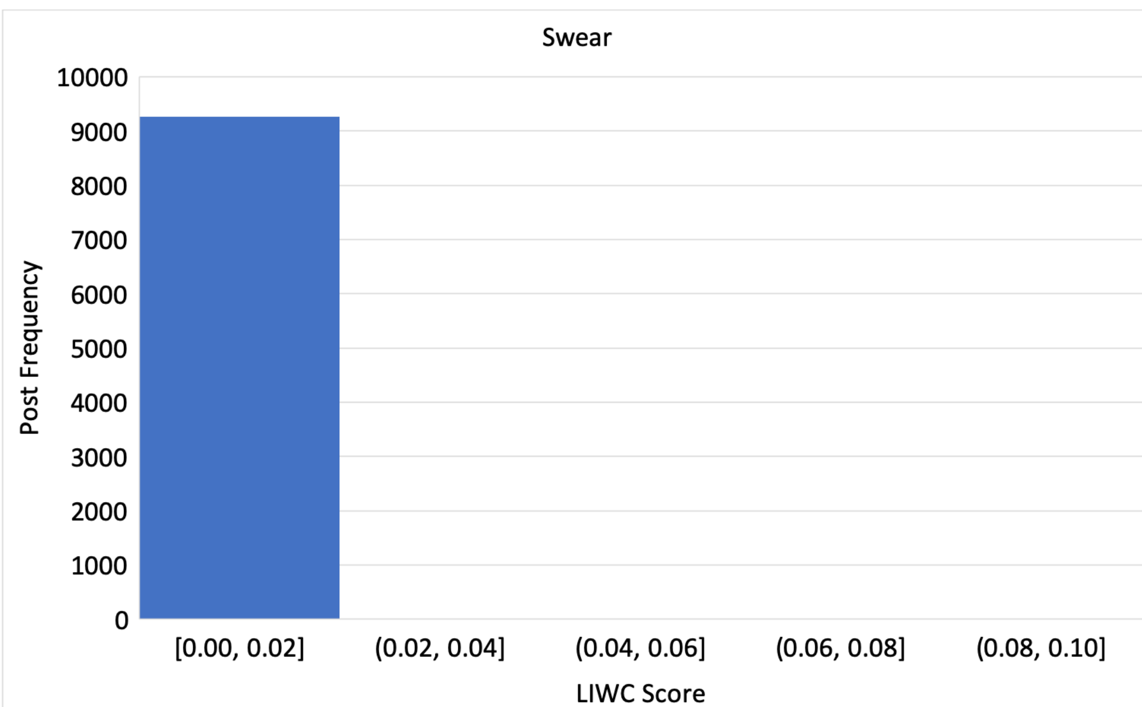


Figure 5.25: The “swear” Category LIWC Score for C4

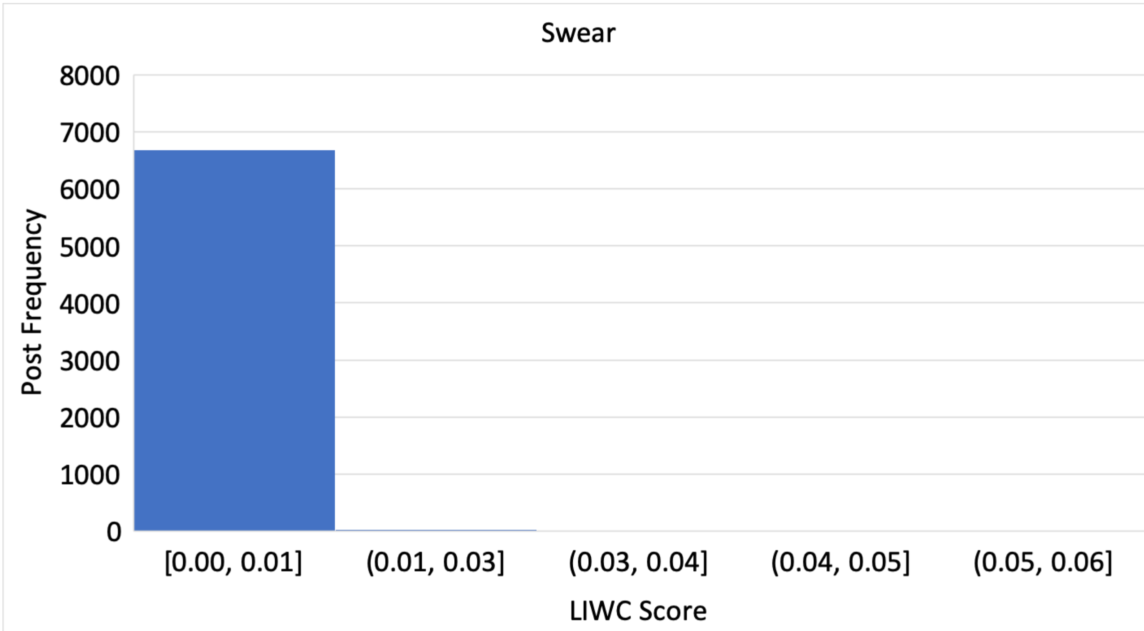


Figure 5.26: The “swear” Category LIWC Score for C5

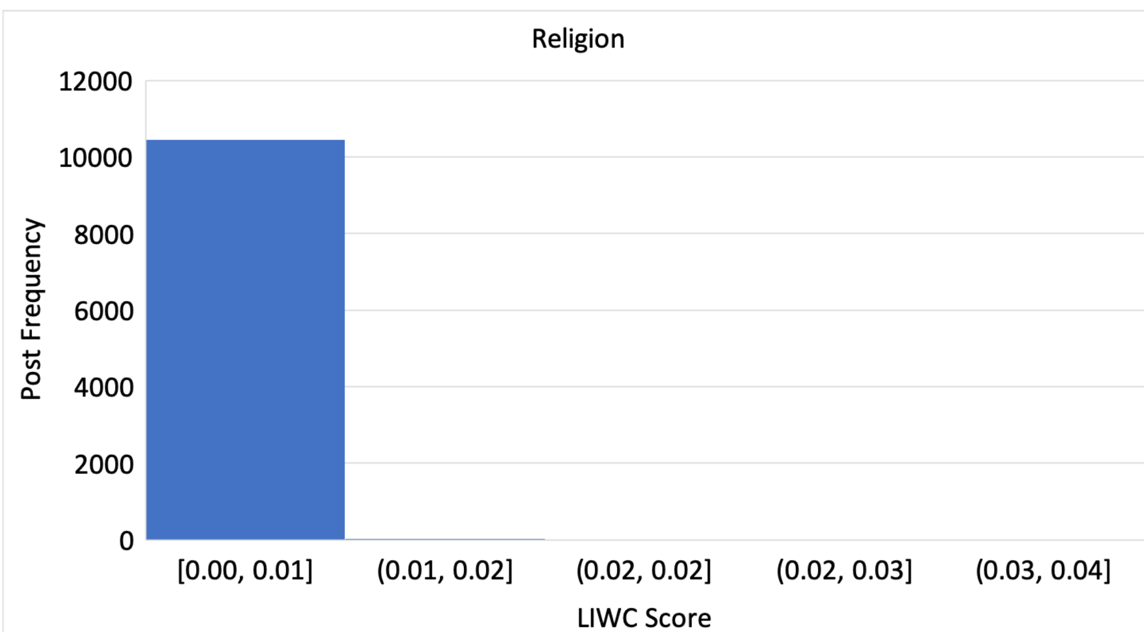


Figure 5.27: The “religion” Category LIWC Score for C1

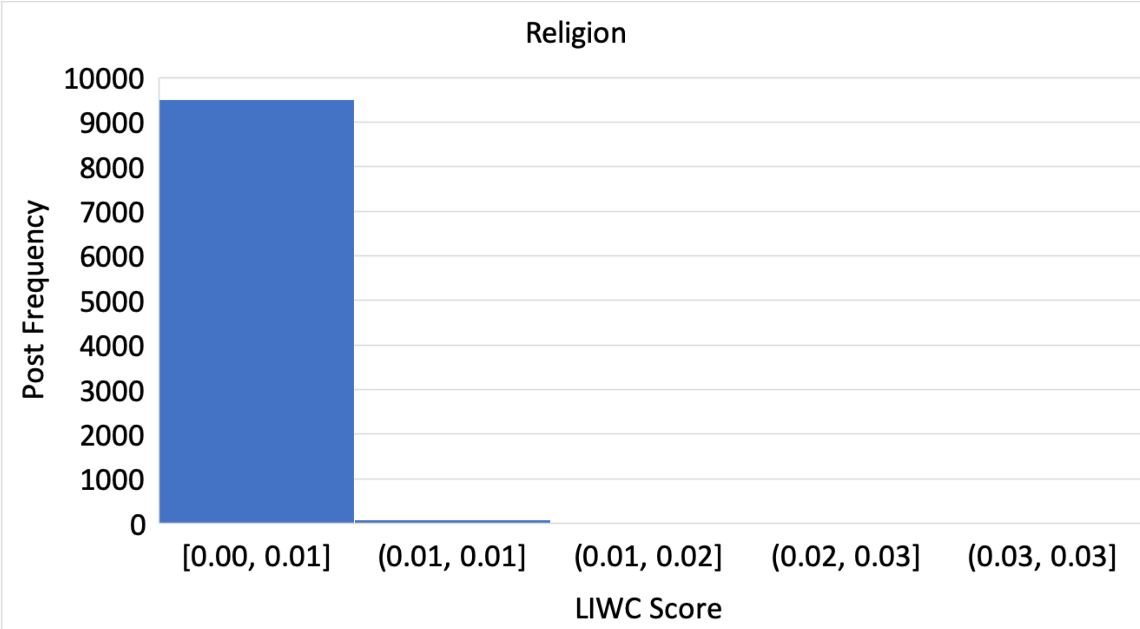


Figure 5.28: The "religion" Category LIWC Score for C2

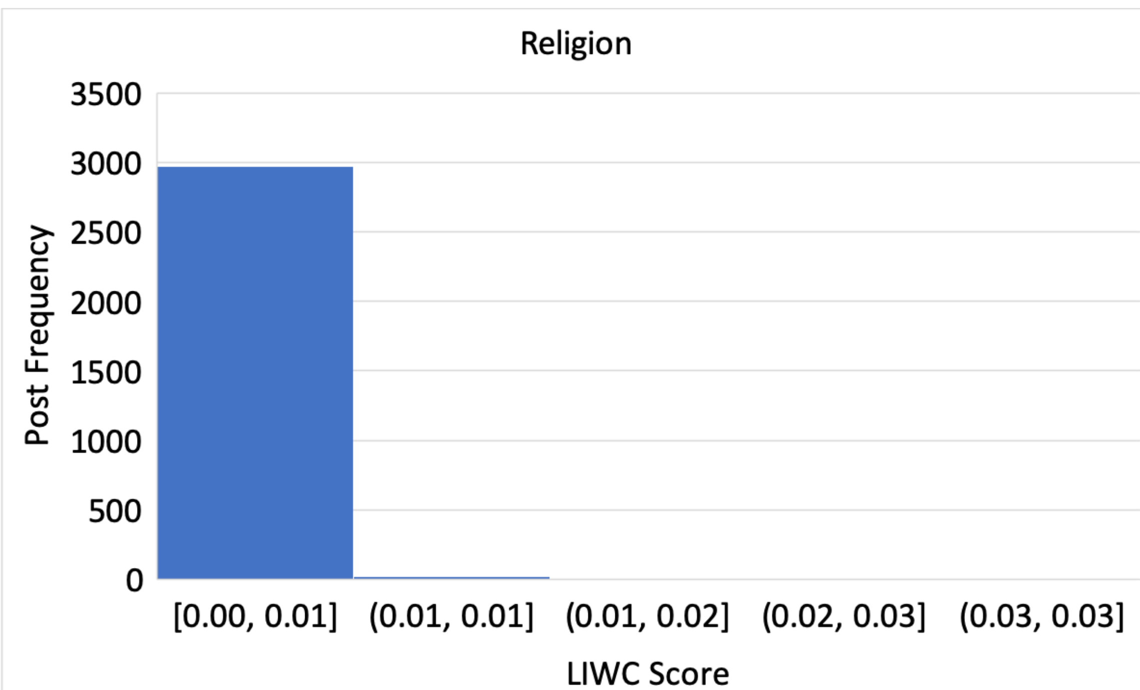


Figure 5.29: The "religion" Category LIWC Score for C3

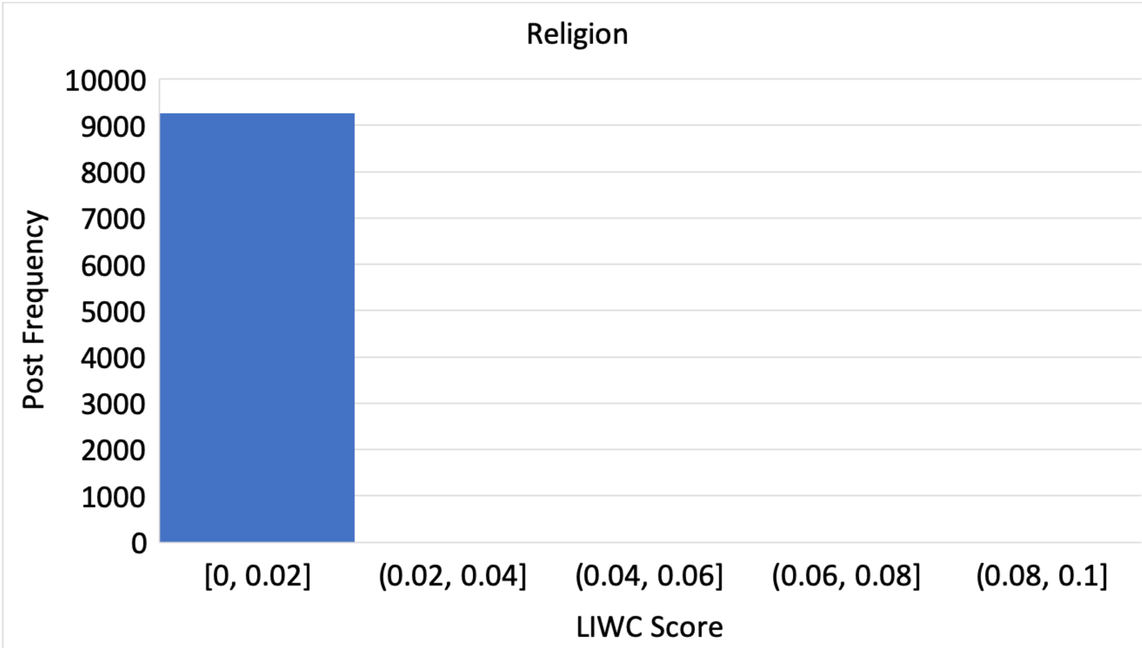


Figure 5.30: The "religion" Category LIWC Score for C4

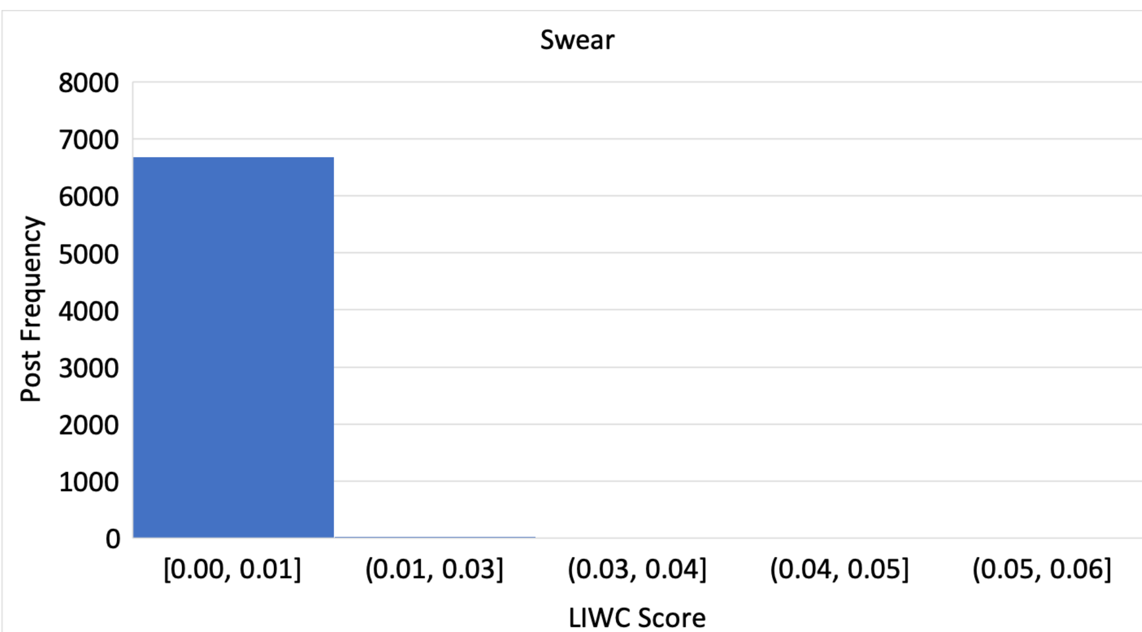


Figure 5.31: The "religion" Category LIWC Score for C5

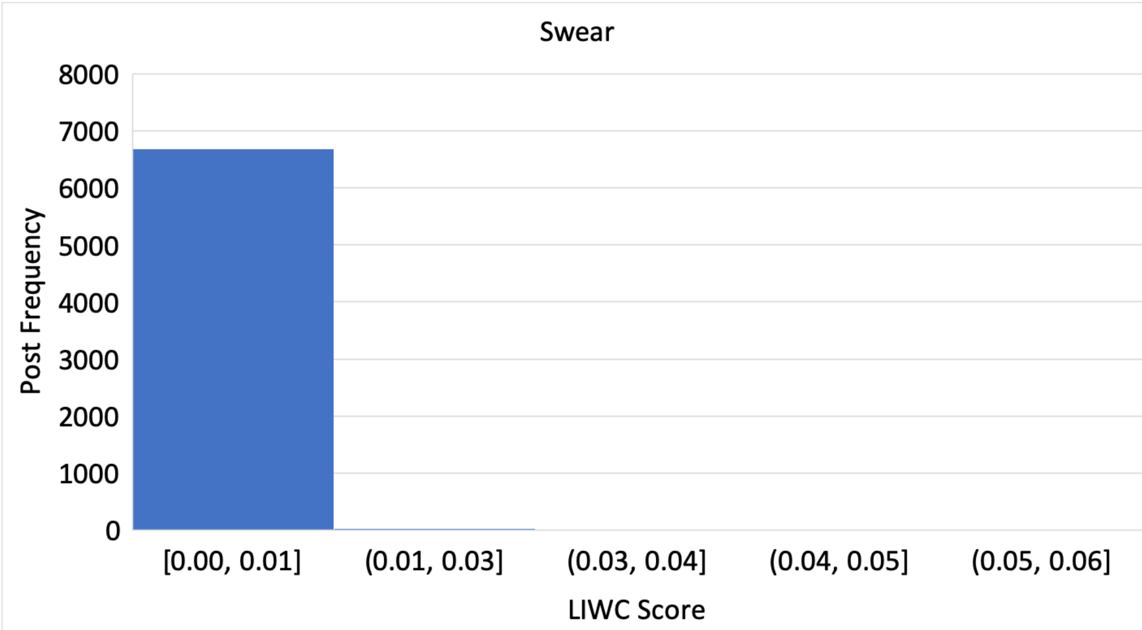


Figure 5.32: The “religion” Category LIWC Score for C5

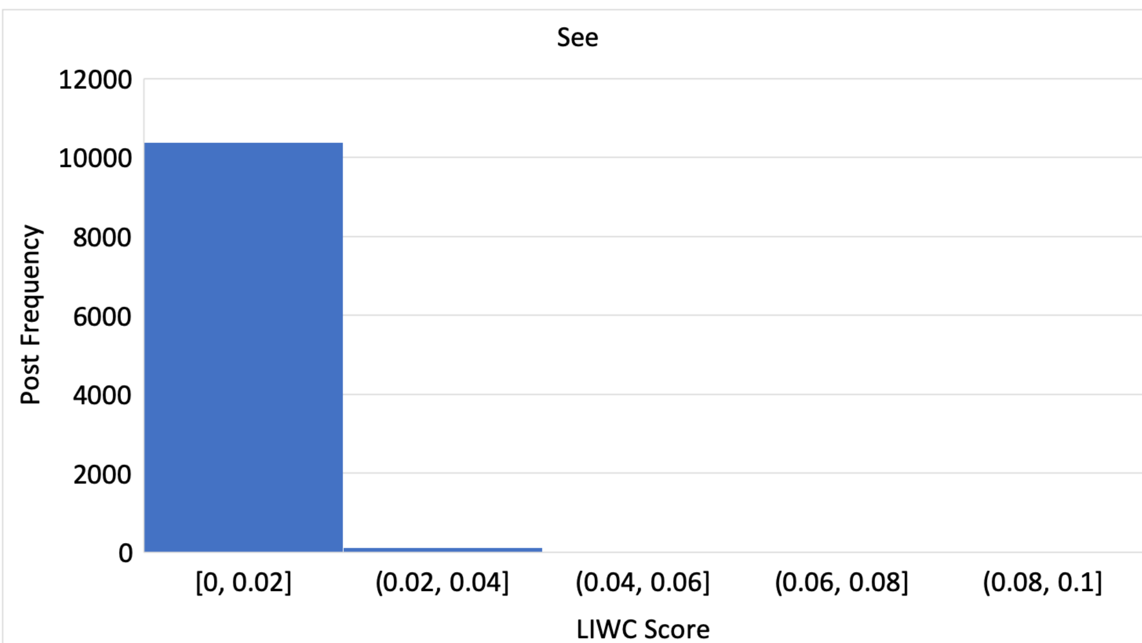


Figure 5.33: The “see” Category LIWC Score for C1

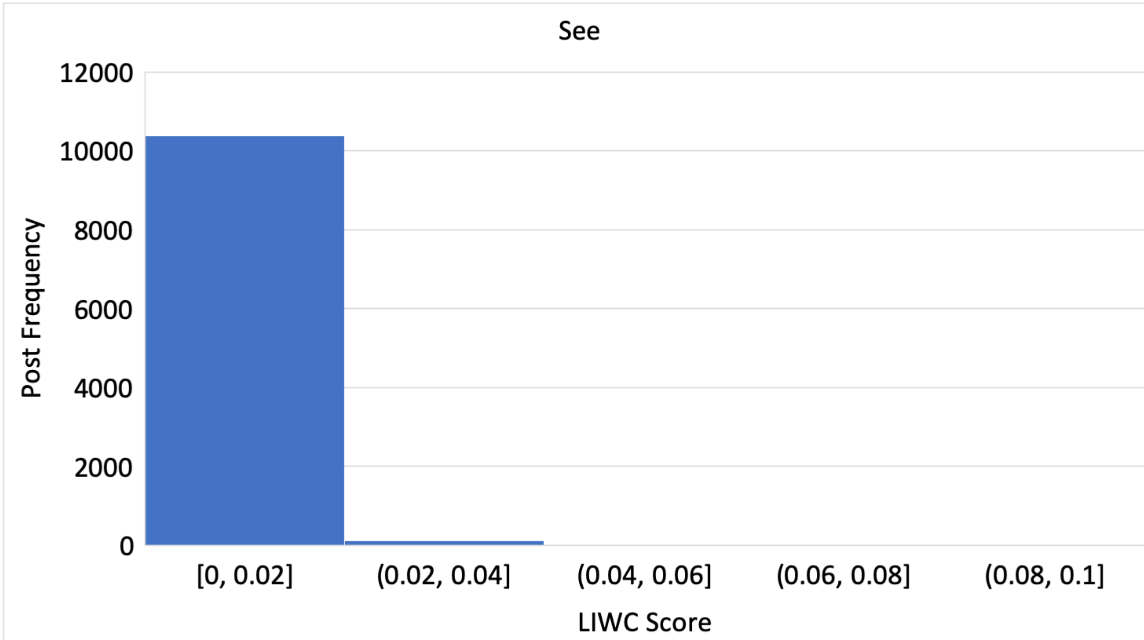


Figure 5.34: The “see” Category LIWC Score for C1

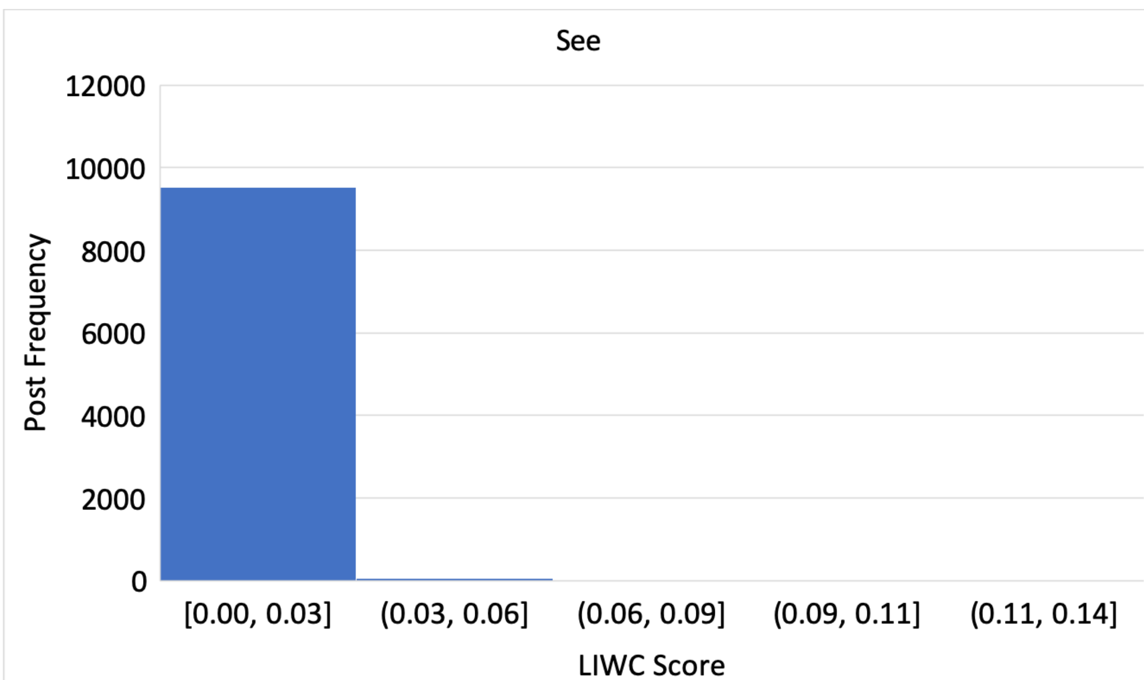


Figure 5.35: The “see” Category LIWC Score for C2

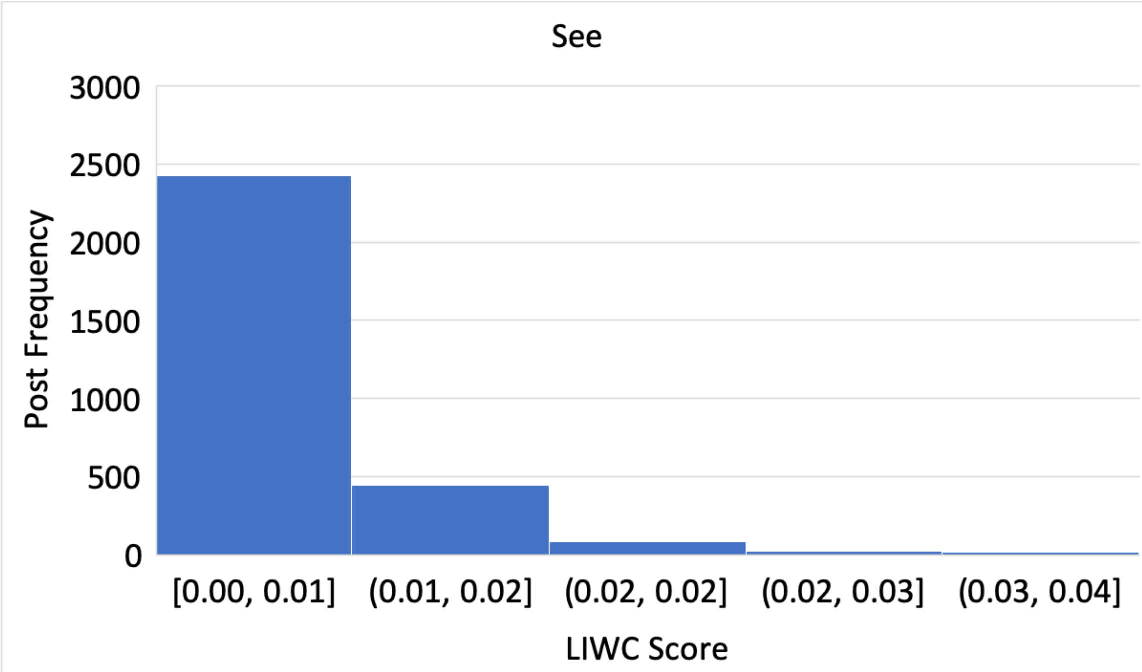


Figure 5.36: The “see” Category LIWC Score for C3

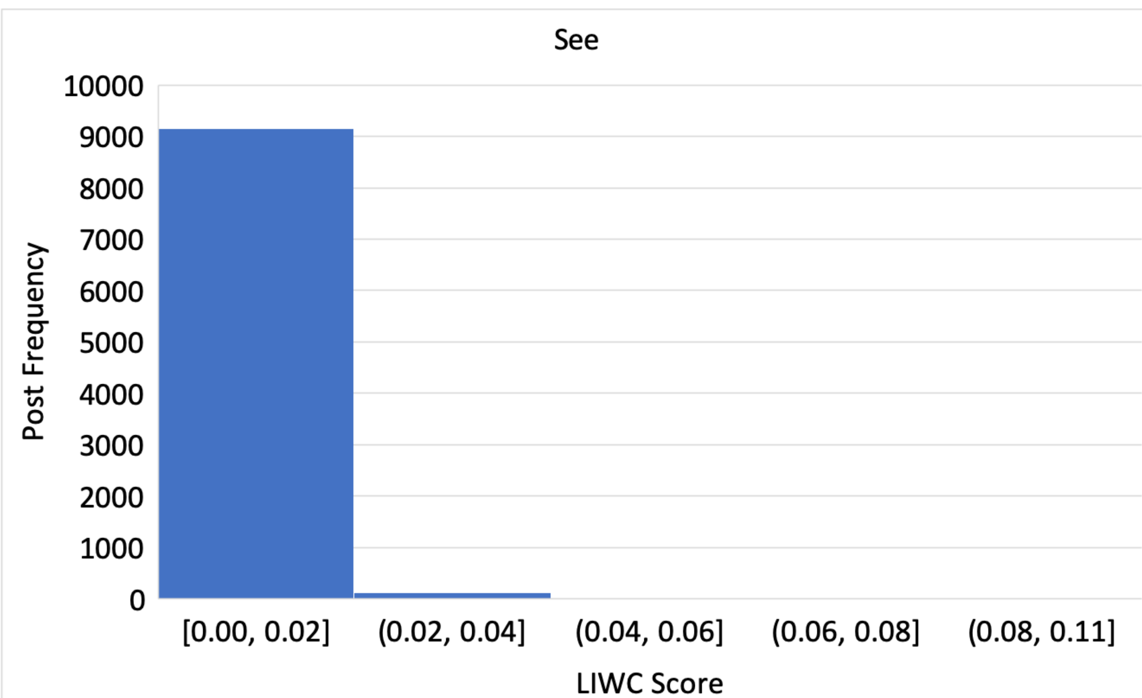


Figure 5.37: The “see” Category LIWC Score for C4

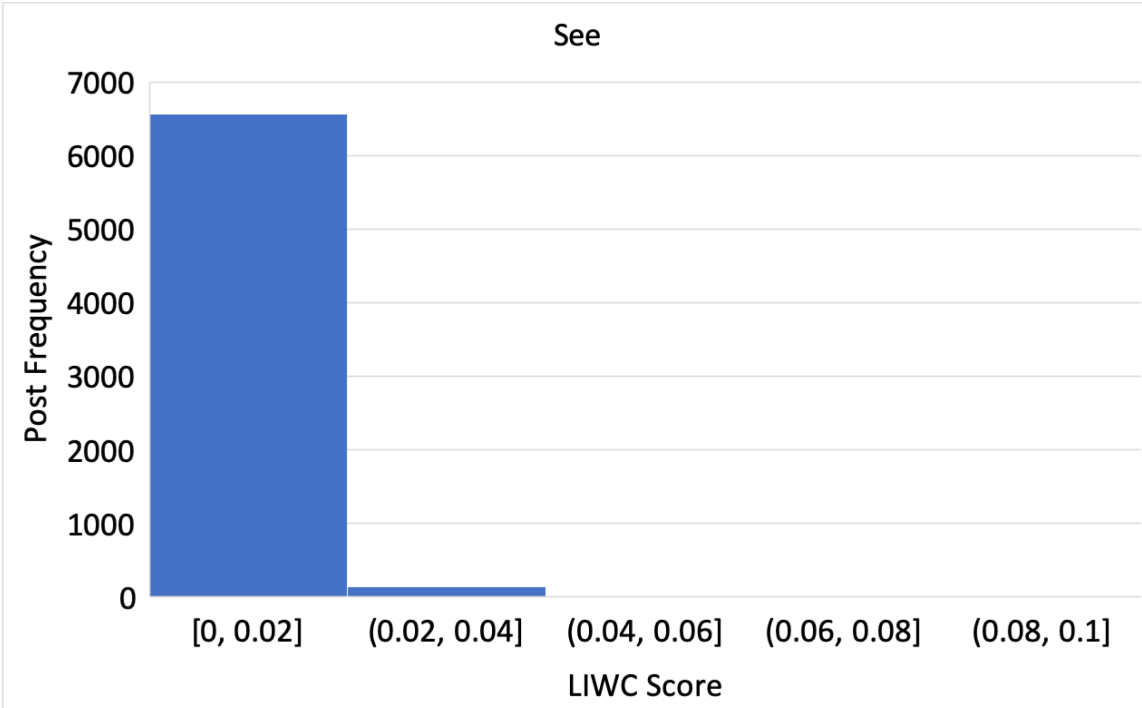


Figure 5.38: The “see” Category LIWC Score for C5

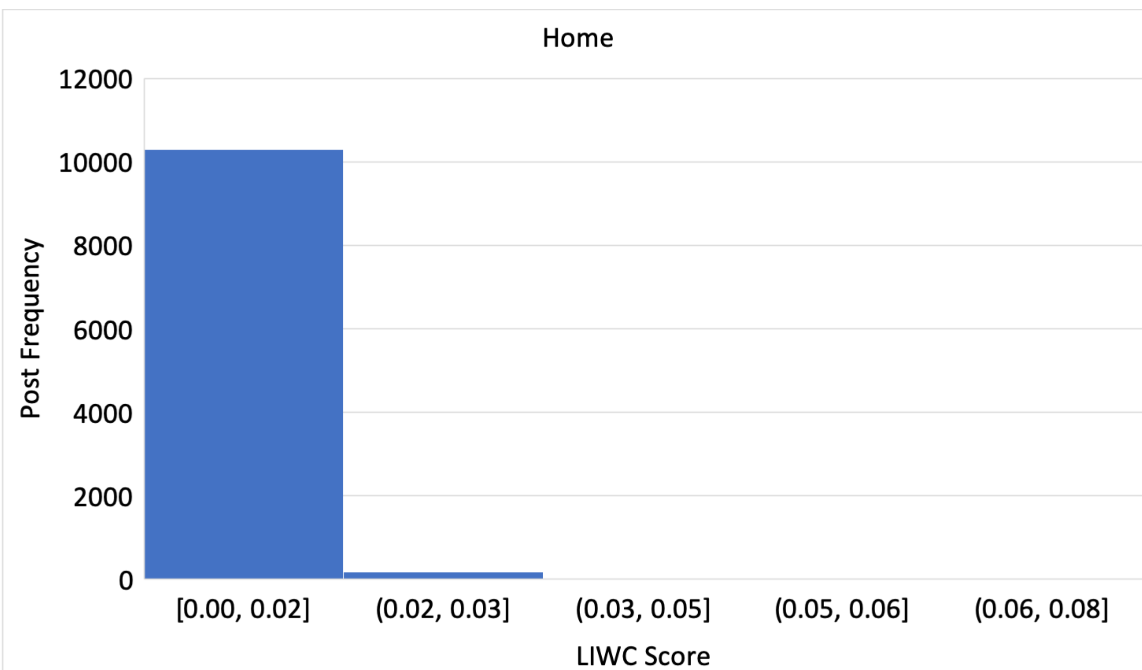


Figure 5.39: The “home” Category LIWC Score for C1

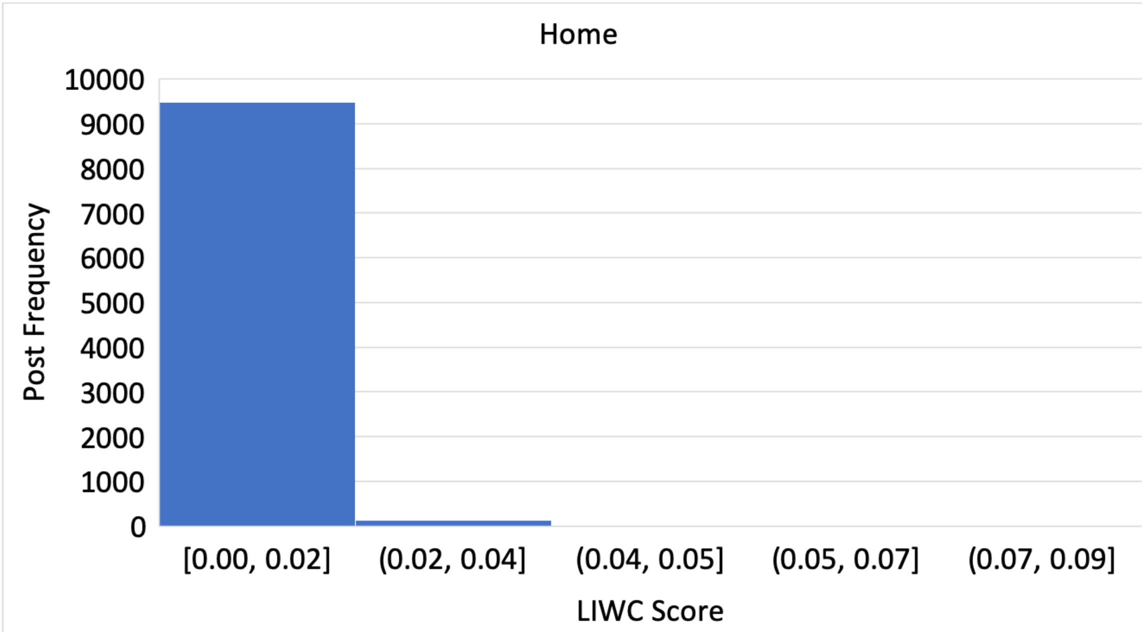


Figure 5.40: The “home” Category LIWC Score for C2

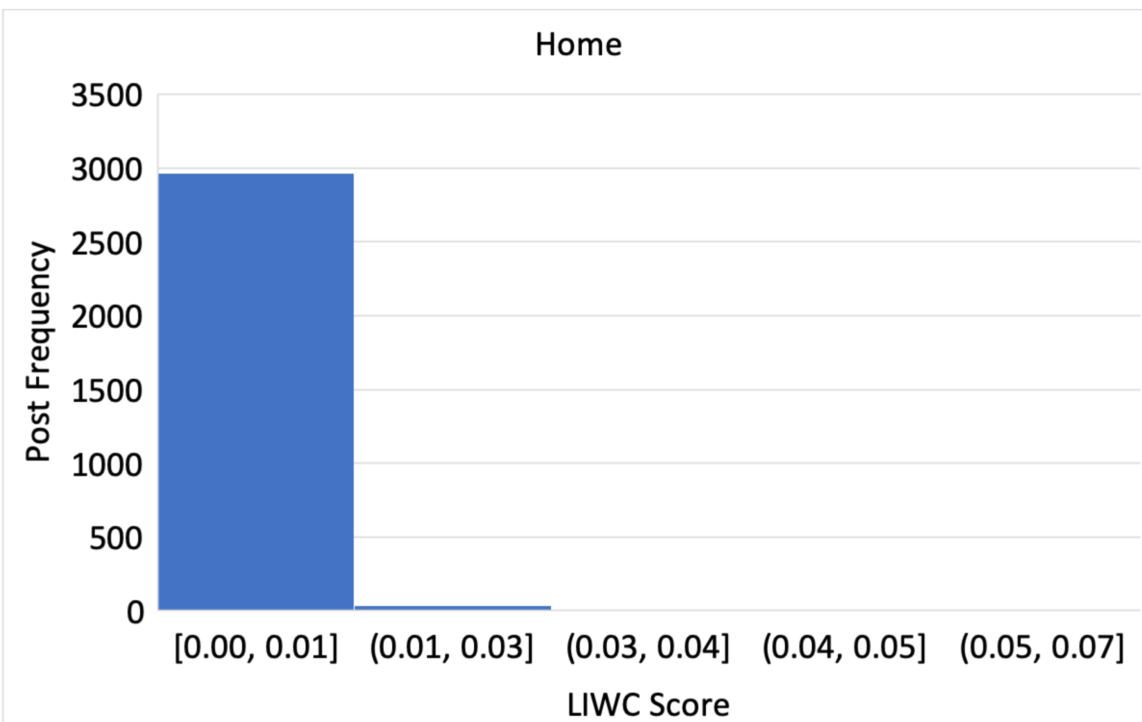


Figure 5.41: The “home” Category LIWC Score for C3

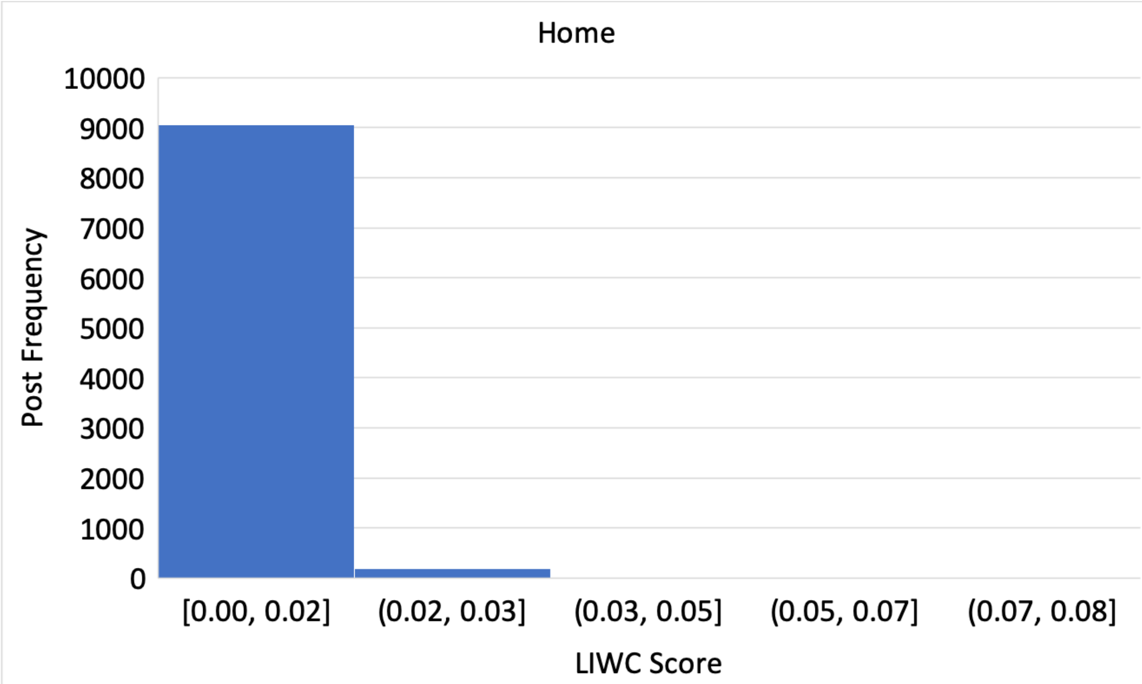


Figure 5.42: The “home” Category LIWC Score for C4

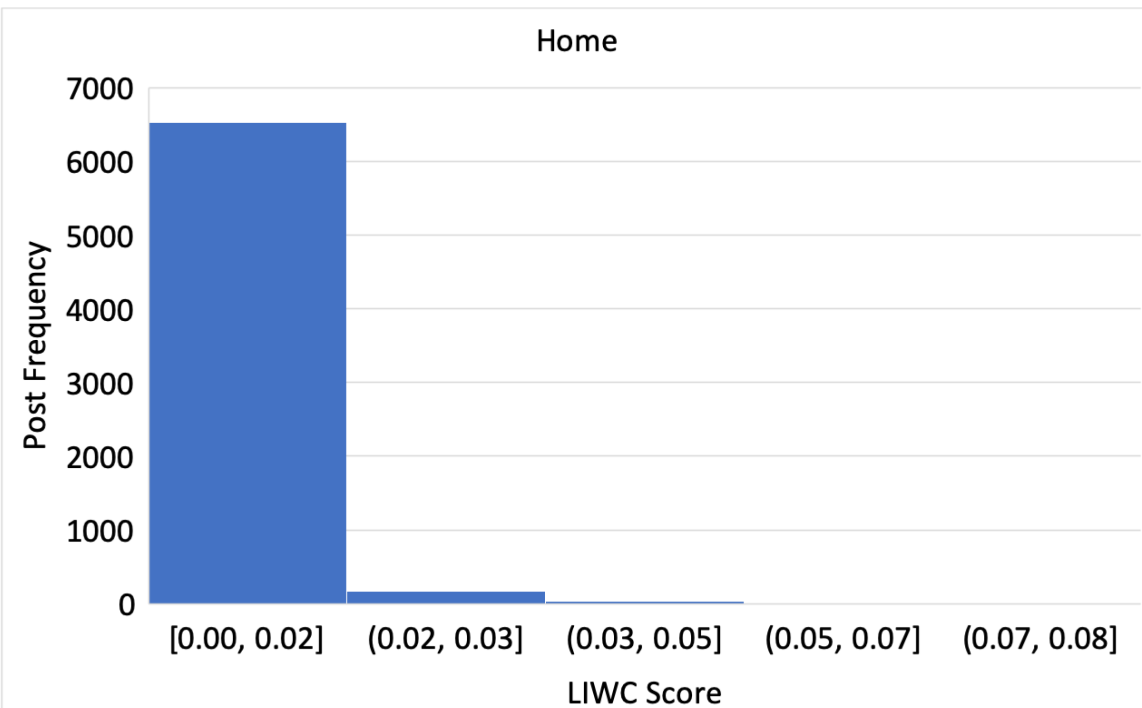


Figure 5.43: The “home” Category LIWC Score for C5

Chapter 6

CONCLUSION

This dissertation investigated the support-seeking as well as support-providing patterns on social media platforms in the context of mental health. We divided the work into two different objectives. The first work focused on the role of social media on mental health specifically among the sexual abuse community. We studied the discussion scope and the linguistic context of online communities related to sexual abuse. Secondly, we extended the work by using the same framework, to study various online mental health communities and the linguistic behaviour of users in these more broad online mental health groups.

We began our first experiment by using social media data from Reddit analyze the linguistic behavior of communication among users. We analyzed the subreddits '/r/rape' and 'r/rapecounseling' to answer the questions surrounding support by extending natural language processing techniques for automated analysis. The analysis first focused on the original posts shared on this subreddit. The second thread of analysis focused on the responses provided to the original posts to understand the behavior of supportive content. It showed that the discussion in the /r/rapecounseling subreddit focused on the action that should be taken by the victim such as *police admit friend* and *support action evidence*. The discussion scope also showed how responses among users indicate that social media is a medium for them to share openly. We found these tri-gram results for /r/rapecounseling highlight the importance of social media as seen in the phrases *public social media* and *rape crisis center*. This finding showed that the users also consider the /r/rapecounseling subreddit as a crisis center for them to refer when it comes to rape. User response revealed the common reluctance to share problems with actual people outside of the Reddit community by mentioning *feel comfortable*, *accusing someone circle*, *never told anyone*, and *selfish people compared*. Third, we ran sentiment analysis on both subreddits to study the sentiment of these two communities. We used the Vader(Valence Aware Dictionary and Sentiment Reasoner) sentiment analysis tool to get the sentiment score for data that we had. Based on the analysis, we found that the criteria of extracted information on two subreddits /r/rape and /r/rapecounseling are different. We found that the results for the /r/rape subreddit were higher in negative score compared to /r/rapecounseling. This finding suggested the /r/rapecounseling subreddit provides less negativity in the discussion than /r/rape and the responses are more likely supportive towards users. This is because /r/rapecounseling includes professionals in its discussion thread which ensure less negative sentiment across the discussion.

For the topic modeling part, we ran LDA on our dataset to extract the top topics for each subreddits. Our LDA results revealed that insurance is also one of the main concerns when it comes to a health-related issues such as getting sick, medical treatment, therapy, etc. Similarly, legal advice is another interesting topic we discovered through our methodology. Note that, some queries for both subreddit were similar and some were different. This is because we only generated the top topic for each subreddit. Legal is one of the top discussed topics in the /r/rape subreddit but it was not a top topic in /r/rapecounseling. Aside from that, we can see in Table 4.7 that users discussed more on trying to recover and finding the right resources for professional help in the /r/rapecounseling subreddit. These results showed that these two subreddits are not exactly similar in terms of topic distribution. Our topic modeling analysis also showed that frequently discussed topics in /r/rape and /r/rapecounseling subreddit are not similar to each other. Finally, sentiment analysis results indicated that /r/rape have higher negative sentiment compared to /r/rapecounseling. This could have been due to the involvement of the professional individuals in the rape counseling subreddit. We were also able to show the differences between /r/rape and /r/rapecounseling subreddits by using a statistical t-test. Results also indicated that these subreddits can be really helpful to a certain group of users and that online communities can be a medium to provide information to support seekers.

For another piece of work in chapter 5, we studied on the sentiment analysis, topic distribution, and emotion attributes across online mental health communities from Reddit. First, we studied how each community is different in terms of its sentiment distribution. We figured that using only post's headlines would not give an accurate sentiment analysis for the community. This means that positive subreddit might contain negative sentiment discussion and vise versa. So, in order to study the sentiment distribution across the mental health community, using the headlines alone was not enough. The headlines can indicate a positive or negative sentiment but the discussion content will have a different distribution. We then further continued our sentiment study by analyzing each comment involved. We found that only Community 1 has higher negative sentiment and the other 4 communities have higher positive sentiments. From the sentiment analysis results, we saw that the distribution of sentiment among online communities was sparse between "very negative" and "very positive". There were very few distributions in the middle of the graph. This was due to the topics discussed in the community. People tend to condemn people or support people depending on the community, which explains the sentiment distribution. This showed that even if the conversation was all about mental health, it does not mean that only negative sentiment is involved. This is because every conversation is different and people nowadays use social media as a medium of information and support-seeking. To support our discussion, we also ran the LDA model on our dataset. The topic distribution of all communities showed various topics across every community. For community 1, users were more focused on discussing their symptoms and seeking information through from Reddit. Moving on to community 2 (C2), topics extracted for this community showed that some group of users discussed on a topic related to mental images. On the other hand, community 3, which represents the compulsive disorder, showed a variation from mental health action. This community talked more on the symptoms of the disorders and on treatment information. Meanwhile, community 4, which represent coping and therapy subreddits, was more likely focused on topics that involve women's health. We can see words like *miscarriage, pregnancy, and husband.* The last community is community 5, which stands for mood disorder and includes eight subreddits. This community was mainly focused on the symptoms, treatment and financial discussion scope. The similarities across all communities are each community will have a topic distribution on social support, feelings, and treatment; which is not surprising. This supported the fact that the user does use an online forum such as Reddit as a medium for them to get support and information especially when it comes to mental health or other related stigmatized issues. We also discussed on how sentiment distribution and topics extracted from each community are correlated to each other. The sentiment distribution for community 1 shows that this community was higher on negative sentiment. At the same time, the topics that were extracted for this community show that users in this community uses the online forum medium to get help and information. They also discussed on their symptoms which can contribute to the higher negative sentiment. Meanwhile, the rest of the communities which are community 2, 3, 4 and 5 have higher positive sentiment. These communities showed that users discuss on a broad range of topics. This contributed to the lower negative sentiment and higher positive sentiment. Based on our observation on both sentiment and topic distribution for all communities, results do correlate and validate our hypothesis. Last, but not least, we also reported LIWC result which represent two categories that we selected based on our previous work. The categories that we chose were “sad”, “swear”, “see”, “home”, “money”, and “religion” as we discussed in chapter 4. These categories rank among the top that can contribute to our understanding of the online mental health communities. From scores reported by each community, we saw various distribution across the communities. This is because each community carries a very different scope of discussion that may lead users to be less or more emotional.

6.1 Future Directions

There are a variety of motivations for users to share specific information on social media: to offer or seek support, to fight the stigma of their illness, or perhaps to offer an explanation for certain behaviors (Coppersmith *et al.*, 2015a). Social media data can benefit health and stigmatized studies. It provides the raw materials for the largest journal study to date, with tens of millions of users and publicly available data. Data gathered via social media is already in digitized form, making it conducive to automated analysis. Existing work showed that it is possible to study stigmatized illnesses by leveraging the large-scale social media data in understanding and analyzing problems and employing machine learning algorithms to understand, measure, and predict stigmatized health problems. More research on social media analysis using machine learning will help advance this important emerging field via multidisciplinary collaboration, research, and development.

Studying mental health issues using social media is challenging. Although a large body of work has emerged in recent years for investigating mental health issues using social media data, there are still open challenges for further investigation. Some potential research directions are suggested below:

- We studied the language of social media from users by leveraging subreddits that share a similar context of the discussion; providing a roadmap for future work. We explored simple techniques capable of distinguishing between mental health subreddits, based on attribute quantified from the users' language. This study can be extended to another health related problem. For example, differentiating between two specific mental health illnesses such as "depression" and "anxiety" by adapting the analysis steps that we used in this dissertation. It can also be elaborated to study the relationship of the owner of the discussion and comments from other users by looking closely at the language attributes from the discussion. Another future direction for this work is studying all the mental health communities by treating them as one big group without separating them based on the categories. We can combine all mental health subreddits and study the correlation of sentiment and topic distribution. Based on this analysis, we can try to detect which subreddits are focusing on a certain topic and what is the sentiment of that particular subreddit. On the other hand, we can also extend the work by extracting additional features for the mental health community such as geographical and temporal features. Once we have this kind of feature, we can do a comparison of mental illnesses in terms of geographical and temporal features. It is also possible to do a comparison between anonymous users and non-anonymous users' activities on online mental health communities.

- The increasing popularity of social media allows users to participate in online activities such as creating online profiles, interacting with other people, expressing opinions and emotions, sharing posts and various personal information. User-generated data on these platforms is rich in content and could reveal information regarding users' mental health situations. However, little attention has been paid on collecting the proper amount of user-information specifically on mental health (Coppersmith *et al.*, 2015c; Nadeem, 2016). One future direction is to collect a proper amount of labeled user-data as a benchmark which requires cooperation between psychologists and computer scientist (Morstatter *et al.*, 2013). This data can include users' behavioral information collected from social media platforms as well as their mental health condition information provided by experts. Preparing such data gives opportunities to both computer scientists and psychologists to benefit from a tremendous amount of data generated in social media platforms to better understand mental health issues and propose solutions to solve them.

- User-generated social media data is heterogeneous and consists of different aspects such as text, image, and link data. Most of the existing work investigates the mental health issues by just incorporating one aspect of social media data. For example, textual information is used in (Park *et al.*, 2012; De Choudhury *et al.*, 2013; De Choudhury and De, 2014; Coppersmith *et al.*, 2015c,b; Nadeem, 2016; Saha and De Choudhury, 2017; Amir *et al.*, 2017; El Sherief *et al.*, 2017), image information is exploited (Andalibi *et al.*, 2015; Kim *et al.*, 2016; Reece and Danforth, 2017), and link data in (Kawachi and Berkman, 2001; De Choudhury *et al.*, 2013) to understand how user-generated information is correlated with people's mental health concerns. One potential research direction is to examine how different combinations of heterogeneous social media data (e.g., a combination of image and link data, combination of textual and link data, etc.) can be utilized to better understand people's behavior and mental health issues concern. Another future direction is to explore how findings from each aspect of social media data are different from each other, (e.g., results with respect to textual data in comparison the findings with respect to link data).

- Most existing work utilizes either human-computer interaction techniques or data mining related techniques. For example, interview and surveys are used to help further study mental health related issues in social media (Burke *et al.*, 2010; Park *et al.*, 2012; De Choudhury *et al.*, 2013; De Choudhury and De, 2014; Kim *et al.*, 2016; Reece and Danforth, 2017). Statistical and computational techniques are leveraged to understand users' behavior with respect to mental health issues (Kawachi and Berkman, 2001; Nadeem, 2016; Saha and De Choudhury, 2017; Amir *et al.*, 2017). However, research can be furthered to exploit both techniques to understand mental health issues in social media (Park *et al.*, 2012; De Choudhury *et al.*, 2013; Reece and Danforth, 2017) and to develop both human-computer interaction and computational techniques specialized for understanding mental health issues for social media data.

Social media data can benefit mental health studies. The existing work shows that it is possible to study mental health by leveraging the large-scale social media data in understanding and analyzing mental health problems and employing machine learning algorithms to understand, measure, and predict mental health problems. More research on social media analysis using machine learning will help advance this important emerging field via multidisciplinary collaboration, research and development.

REFERENCES

Abbasi, A., H. Chen and A. Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums", ACM Transactions on Information Systems (TOIS) **26**, 3, 12 (2008).

Ajzen, I., "The theory of planned behavior", Organizational behavior and human decision processes **50**, 2, 179–211 (1991).

Amir, S., G. Coppersmith, P. Carvalho, M. J. Silva and B. C. Wallace, "Quantifying mental health from social media with neural user embeddings", arXiv preprint arXiv:1705.00335 (2017).

Andalibi, N., O. L. Haimson, M. De Choudhury and A. Forte, "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", pp. 3906–3918 (ACM, 2016).

Andalibi, N., P. Ozturk and A. Forte, "Depression-related imagery on instagram", in "Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative work & social computing", pp. 231–234 (ACM, 2015).

Astudillo, R., S. Amir, W. Ling, M. Silva and I. Trancoso, "Learning word representations from scarce and noisy data with embedding subspaces", in "Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)", vol. 1, pp. 1074–1084 (2015).

Baker, D. and S. Fortune, "Understanding self-harm and suicide websites: a qualitative interview study of young adult website users", Crisis **29**, 3, 118–122 (2008).

Bay, H., T. Tuytelaars and L. Van Gool, "Surf: Speeded up robust features", in "European conference on computer vision", pp. 404–417 (Springer, 2006).

Blei, D. M., A. Y. Ng and M. I. Jordan, "Latent dirichlet allocation", Journal of machine Learning research **3**, Jan, 993–1022 (2003).

Bourgeault, I., R. Dingwall and R. De Vries, *The SAGE handbook of qualitative methods in health research* (Sage, 2010).

Braithwaite, S. R., C. Giraud-Carrier, J. West, M. D. Barnes and C. L. Hanson, "Validating machine learning algorithms for twitter data against established measures of suicidality", JMIR mental health **3**, 2 (2016).

Burke, M., C. Marlow and T. Lento, "Social network activity and social well-being", in "Proceedings of the SIGCHI conference on human factors in computing systems", pp. 1909–1912 (ACM, 2010).

Chancellor, S., Z. Lin, E. L. Goodman, S. Zerwas and M. De Choudhury, "Quantifying and predicting mental illness severity in online pro-eating disorder communities", in "Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing", pp. 1171–1184 (ACM, 2016).

Cohan, A., S. Young, A. Yates and N. Goharian, "Triaging content severity in online mental health forums", Journal of the Association for Information Science and Technology **68**, 11, 2675–2689 (2017).

Coppersmith, G., M. Dredze, C. Harman and K. Hollingshead, "From adhd to sad: Analyzing the language of mental health on twitter through self-reported diagnoses", in "Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality", pp. 1–10 (2015a).

Coppersmith, G., M. Dredze, C. Harman, K. Hollingshead and M. Mitchell, "Clpsych 2015 shared task: Depression and ptsd on twitter", in "Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality", pp. 31–39 (2015b).

Coppersmith, G., R. Leary, E. Whyne and T. Wood, "Quantifying suicidal ideation via language usage on social media", in "Joint Statistics Meetings Proceedings, Statistical Computing Section, JSM", (2015c).

Daume III, H., "Frustratingly easy domain adaptation", arXiv preprint arXiv:0907.1815 (2009).

De Choudhury, M., S. Counts, E. J. Horvitz and A. Hoff, "Characterizing and predicting postpartum depression from shared facebook data", in "Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing", pp. 626–638 (ACM, 2014).

De Choudhury, M. and S. De, "Mental health discourse on reddit: Self-disclosure, social support, and anonymity.", in "ICWSM", (2014).

De Choudhury, M., M. Gamon, S. Counts and E. Horvitz, "Predicting depression via social media.", ICWSM **13**, 1–10 (2013).

De Choudhury, M. and E. Kiciman, "The language of social support in social media and its effect on suicidal ideation risk", in "Proceedings of the... International AAAI Conference on Weblogs and Social Media. International AAAI Conference on Weblogs and Social Media", vol. 2017, p. 32 (NIH Public Access, 2017).

De Choudhury, M., E. Kiciman, M. Dredze, G. Coppersmith and M. Kumar, "Discovering shifts to suicidal ideation from mental health content in social media", in "Proceedings of the 2016 CHI conference on human factors in computing systems", pp. 2098–2110 (ACM, 2016).

El Sherief, M., E. M. Belding and D. Nguyen, "# notokay: Understanding gender-based violence in social media.", in "ICWSM", pp. 52–61 (2017).

Farnham, S. D. and E. F. Churchill, “Faceted identity, faceted lives: social and technical issues with being yourself online”, in “Proceedings of the ACM 2011 conference on Computer supported cooperative work”, pp. 359–368 (ACM, 2011).

Gilbert, C. H. E., “Vader: A parsimonious rule-based model for sentiment analysis of social media text”, in “Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf”, (2014).

Grieve, R., M. Indian, K. Witteveen, G. A. Tolan and J. Marrington, “Face-to-face or facebook: Can social connectedness be derived online?”, Computers in Human Behavior **29**, 3, 604–609 (2013).

Harman, G. and M. H. Dredze, “Measuring post traumatic stress disorder in twitter”, In ICWSM (2014).

Homan, C. M., N. Lu, X. Tu, M. C. Lytle and V. Silenzio, “Social structure and depression in trevorspace”, in “Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing”, pp. 615–625 (ACM, 2014).

Kamarudin, N. S., V. Rakesh, G. Beigi, L. Manikouda and H. Liu, “A study of reddit-user’s response to rape”, in “2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)”, pp. 591–592 (IEEE, 2018).

Kawachi, I. and L. F. Berkman, “Social ties and mental health”, Journal of Urban health **78**, 3, 458–467 (2001).

Kim, E., J.-A. Lee, Y. Sung and S. M. Choi, “Predicting selfie-posting behavior on social networking sites: An extension of theory of planned behavior”, Computers in Human Behavior **62**, 116–123 (2016).

Kumar, M., M. Dredze, G. Coppersmith and M. De Choudhury, “Detecting changes in suicide content manifested in social media following celebrity suicides”, in “Proceedings of the 26th ACM Conference on Hypertext & Social Media”, pp. 85–94 (ACM, 2015).

Lin, Z., N. Salehi, B. Yao, Y. Chen and M. S. Bernstein, “Better when it was smaller? community content and behavior after massive growth.”, in “ICWSM”, pp. 132–141 (2017).

Ma, X., J. Hancock and M. Naaman, “Anonymity, intimacy and self-disclosure in social media”, in “Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems”, pp. 3857–3869 (ACM, 2016).

Manikonda, L., G. Beigi, H. Liu and S. Kambhampati, “Twitter for sparking a movement, reddit for sharing the moment:# metoo through the lens of social media”, arXiv preprint arXiv:1803.08022 (2018).

Manikonda, L. and M. De Choudhury, “Modeling and understanding visual attributes of mental health disclosures in social media”, in “Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems”, pp. 170–181 (ACM, 2017).

Mitchell, M., K. Hollingshead and G. Coppersmith, “Quantifying the language of schizophrenia in social media”, in “Proceedings of the 2nd workshop on Computational linguistics and clinical psychology: From linguistic signal to clinical reality”, pp. 11–20 (2015).

Morstatter, F., J. Pfeffer, H. Liu and K. M. Carley, “Is the sample good enough? comparing data from twitter’s streaming api with twitter’s firehose”, in “Seventh international AAAI conference on weblogs and social media”, (2013).

Nadeem, M., “Identifying depression on twitter”, arXiv preprint arXiv:1607.07384 (2016).

News, U., “Instagram rated worst social network for mental health”, (2017).

Pan, S. J. and Q. Yang, “A survey on transfer learning”, IEEE Transactions on knowledge and data engineering **22**, 10, 1345–1359 (2010).

Park, M., C. Cha and M. Cha, “Depressive moods of users portrayed in twitter”, in “Proceedings of the ACM SIGKDD Workshop on healthcare informatics (HI-KDD)”, vol. 2012, pp. 1–8 (ACM New York, NY, 2012).

Pedersen, T., “Screening twitter users for depression and ptsd with lexical decision lists”, in “Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality”, pp. 46–53 (2015).

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, “Scikit-learn: Machine learning in python”, Journal of machine learning research **12**, Oct, 2825–2830 (2011).

Peng, X., L.-K. Chi and J. Luo, “The effect of pets on happiness: A large-scale multi-factor analysis using social multimedia”, **9**, 4 (2005).

Pennebaker, J. W., R. L. Boyd, K. Jordan and K. Blackburn, “The development and psychometric properties of liwc2015”, Tech. rep. (2015).

Purohit, H., T. Banerjee, A. Hampton, V. L. Shalin, N. Bhandutia and A. P. Sheth, “Gender-based violence in 140 characters or fewer: A bigdata case study of twitter”, arXiv preprint arXiv:1503.02086 (2015).

Reagan, A. J., C. M. Danforth, B. Tivnan, J. R. Williams and P. S. Dodds, “Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs”, EPJ Data Science **6**, 1, 28 (2017).

Reece, A. G. and C. M. Danforth, “Instagram photos reveal predictive markers of depression”, EPJ Data Science **6**, 1, 15 (2017).

Robinson, J., E. Bailey and S. Byrne, "Social media can be bad for youth mental health, but there are ways it can help", (2017).

Rude, S., E.-M. Gortner and J. Pennebaker, "Language use of depressed and depression-vulnerable college students", Cognition & Emotion **18**, 8, 1121–1133 (2004).

Sable, M. R., F. Danis, D. L. Mauzy and S. K. Gallagher, "Barriers to reporting sexual assault for women and men: Perspectives of college students", Journal of American College Health **55**, 3, 157–162 (2006).

Saha, K. and M. De Choudhury, "Modeling stress with social media around incidents of gun violence on college campuses", (2017).

Sharma, E. and M. De Choudhury, "Mental health support and its relationship to linguistic accommodation in online communities", in "Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems", p. 641 (ACM, 2018).

Shickel, B., M. Heesacker, S. Benton and P. Rashidi, "Hashtag healthcare: From tweets to mental health journals using deep transfer learning", arXiv preprint arXiv:1708.01372 (2017).

Tsugawa, S., Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh and H. Ohsaki, "Recognizing depression from twitter activity", in "Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems", pp. 3187–3196 (ACM, 2015).

Ullman, S. E. and H. H. Filipas, "Predictors of ptsd symptom severity and social reactions in sexual assault victims", Journal of traumatic stress **14**, 2, 369–389 (2001).

Van House, N. A. and M. Davis, "The social life of cameraphone images", in "Proceedings of the Pervasive Image Capture and Sharing: New Social Practices and Implications for Technology Workshop (PICS 2005) at the Seventh International Conference on Ubiquitous Computing (UbiComp 2005)", (Citeseer, 2005).

APPENDIX A

ACCOMPLISHMENTS AND PUBLICATIONS

A.1 ACADEMIC ACHIEVEMENTS

- Awarded PhD Scholarship by Ministry of Higher Education Malaysia and University Malaysia Pahang from 2015 to 2019
- The Best Book Award for Science and Technology Category, Terengganu, Malaysia Book Title: A Practical Approach for Image Classification 2016
- Pulse of The Nation Award by Malaysia Government 2011 for outstanding performance in academic and leadership
- ASU Travel Award to CVPR 2016 in Las Vegas, Nevada
- ASU Travel Award to ASONAM 2018 in Barcelona, Spain
- Invited Book Chapter Reviewer: Springer Book Chapter
- Invited Journal Reviewer: IEEE Intelligent Systems, SNAM
- Invited Conference Reviewer: SIGIR,IJCAI,ICWSM,WWW,WSDM, AAAI,ASONAM,ICDM,SDM, SBP-BRIM, MICCAI, CVPR

A.2 PUBLICATIONS

A.2.1 CONFERENCE PAPERS

- Kamarudin, N.S, Mohan, V., Beigi, Gh., Manikonda, L., Liu, H. A Study of Reddit- User's Response to Rape. Presented at the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM-18)

- Kamarudin, N.S, Makhtar, M., Syed Abdullah, F, (2015). Shape-Based Single Object Classification using Ensemble Method Classifiers. In UniSZA Research Conference.

- Kamarudin, N.S., Makhtar, M., Fadzli, E. (2013). Collection of Amazon Dataset Towards Automatic Image Annotation. In UniSZA Postgraduate Conference 2013 (pp. 1–7). University Sultan Zainal Abidin

- Kamarudin, N.S., Beigi, Gh., L., Liu, H.: A Study on Mental Health Discussion through Reddit'. Submitted to WWW 2020.

A.2.2 JOURNAL PAPERS

- Kamarudin, N.S, Beigi, Gh., Mohan, V., Manikonda, L., Liu, H.: 'Nuggets Discovered from User Response to Rape Discussions through Reddit'. To be Submitted
- Kamarudin, S., Makhtar, M., Syed Abdullah, F., Mohamad, F. S., Mohamad, M., Kadir, M. F. A. (2015). Comparison Of Image Classification Techniques. Journal of Theoretical and Applied Information Technology, vo 71, no.1
- Makhtar, M., Syed Abdullah, F., Kamarudin, N.S, Mohamad, F. S., Abdul Kadir, M. F., Mohamad, M. (2014). The Contribution of Feature Selection and Morphological Operation For On- Line Business System's Image Classification. World Applied Science Journal, 10(11), 303–314.

A.2.3 BOOK CHAPTER

- Kamarudin, N.S, Beigi, Gh., Manikonda, L., Liu, H.: 'Social Media– A New and Effective Platform for Raising Awareness of Mental Health Issues'. Accepted for inclusion in the edited volume entitled as Open Source Intelligence and Security Informatics. Published by Springer.

A.2.4 THESES AND BOOK

- Makhtar, M., Syed Abdullah, F., Kamarudin, N.S. A Practical Approach for Image Classification. Published (2015).
- Kamarudin, N.S, Makhtar, M., Syed Abdullah, F., Framework for Single Object Images in Multiple Categories using Ensemble Method. M.Sc Thesis (2014).
- Kamarudin, N.S, Syed Abdullah, F., UniSZA Malay Word Stemmer Algorithm. B.ScThesis, 2012.