Understanding the Societal Disruption due to COVID-19 via User Tweets

Swaroop Gowdra Shanthakumar, Anand Seetharam, Arti Ramesh
Computer Science Department, SUNY Binghamton
(sgowdra1, aseethar, artir)@binghamton.edu

***Abstract*—In this paper, we collect data from Twitter and conduct a linguistic analysis of the user tweets to understand the social and economic disruption caused by the COVID-19 pandemic. To better appreciate peoples' opinions and concerns with regards to the socio-economic conditions of *addiction*, *mental health*, *unemployment* and *immigration*, we collect data for a period of approximately 3 months in the beginning of the pandemic. We analyze the term and co-occurrence frequencies to identify the most commonly occurring words and bigrams in the discussion for each of the four categories. We then conduct semantic role labeling to determine the action words in each category and then adopt a LSTM-based dependency parsing model to identify the main nouns linked with these action words. We then adopt a seeded topic modeling approach to automatically identify the main topics of discussion in each category. We finally conclude with a sentiment analysis of the tweets in each category to determine the overall sentiment associated with each category. Our fine-grained linguistic study unearths the difficulties experienced by the people (e.g., action verb *need* associated with nouns such as *aid* and *assistance* in the unemployment category). We also observe that the overall sentiment in the tweets is negative, driven by people experiencing the pains of job loss, deportation, and the difficulty in accessing programs and treatments related to addiction. Our analysis highlights the main challenges experienced by the people during the start of the COVID-19 crisis and lays the foundation for recognizing and developing the most pertinent public and social policies so as to minimize peoples' suffering in case of a future pandemic.**

I. Introduction

The COVID-19 pandemic has caused immense grief and suffering to people around the world. The lockdowns, physical distancing and quarantining measures necessary to curb the spread of the disease have resulted in severe disruption to social life and have adversely impacted the employment and mental well-being of millions of individuals. Therefore, this paper aims to understand and quantify the societal impact of the current COVID-19 pandemic so that the learnings will help us better prepare for future pandemics.

In this paper, we investigate the societal impact of COVID-19 by collecting user tweets from Twitter, one of the most popular social media platforms for people to express their opinions. Mining user tweets enables us to conduct a large-scale study and helps us understand user opinions as expressed by them. Our goal here is to conduct an in-depth linguistic study to specifically investigate the impact of the pandemic on socio-economic conditions such as addiction, mental health, unemployment and immigration. We build on our prior research [1], where we collect and analyze peoples' reaction to the COVID-19 pandemic and containment measures such as lockdowns and school closures during its early days.

We collect approximately 50,000 tweets that explicitly mention COVID-19 and also contain keywords related to *addiction*, *mental health*, *unemployment* and *immigration* to investigate the unfortunate impact of COVID-19 on these issues. We begin our linguistic analysis by determining the most frequent words in each of these categories and then extract bigrams to identify meaningful co-occurrence of words. For example, in the unemployment category, we observe top words such as *'claim'* and *'work'*, and bigrams such as *'unemployment claim'* and *'million americans'* that underscores the plight of millions of individuals because of job loss.

We perform semantic role labeling to determine the most frequent action words. For example, we observe verbs such as *'struggle'*, *'stop'* and *'deal'* in the addiction category that point to the major struggles encountered by the people. Similarly, in the unemployment category, we observe verbs such as *'need'*, *'work'* and *'lose'*, which highlights the hardships experienced by people due to job loss. We adopt a LSTM-based dependency parser to perform a contextual analysis of these action words and determine the key nouns associated with them. For example, in the addiction category we observe that the action word *'stop'* contextually occurs with nouns such as *coverage* and *program*, which indicates that people are worried about their insurance coverage and some support programs getting stopped due to pandemic.

We design a seeded topic modeling approach (*seeded LDA*) to automatically identify the main topics of discussion in each category (e.g., addiction) and to categorize the tweets to these topics. We identify 2 main topics of discussion in the addiction and immigration categories, and 3 main topics in the mental health and unemployment categories. Additionally, for all categories except immigration, we find that there is a need for a miscellaneous topic to group the remaining tweets. For example in the addiction category, we identify two topics—*'mental state'* and *'struggles'* apart from a miscellaneous topic. In the *'mental state'* topic, we observe people discussing about serious mental health issues such as depression, anxiety and stress, while in the *'struggles'* topic, we observe the presence of words such as difficult and cope.

Finally, we conduct sentiment analysis to unearth the underlying sentiment in the tweets. We observe that the overall sentiment related to addiction, unemployment and immigration is highly negative. Given the large number of layoffs, business closures and lockdowns, it is understandable that the majority of tweets in the unemployment category echo a negative sentiment. Interestingly, we observe that a larger percentage of tweets in the mental health category contain a positive sentiment, primarily driven by tweets motivating people to take care of their mental well-being. In the mental health category, we observe words such as *'improve'*, *'manage'* and *'cope'*, which helps us interpret the higher positive sentiment in this category. Our analysis highlights the effect of the pandemic on deep societal issues such as addiction, mental health, unemployment and immigration and paves the way for more fine-grained analysis in future.

II. Related Work

In this section, we outline research related to analyzing Twitter and other social media data to understand social, political and economic impacts of a variety of different events.

Ahmed et al. focus on the conspiracy theories surrounding the novel coronavirus, especially in relationship with 5G [2]. Other work in this area include analyzing Twitter communications to understand the possibility of using bots for propagating misinformation and political conspiracies during the pandemic [3], [4]. The authors conduct infodemiology studies on Twitter communications to understand how information is spreading during this time in South Korea [5]. The social stigma and misinformation created by referencing the novel coronavirus as the 'Chinese virus' is investigated in [6]. Xue et al. examine COVID-19–related discussions, concerns, and sentiments using tweets posted by Twitter users [7].

Twitter communications and social media data have long been used to understand public opinion surrounding many important societal problems. Some of the pioneering work involving social media data are Perez et al.'s work on understanding symptoms of irritable bowel disease using Twitter communications [8]. Dredze et al. [9] and Paul et al. [10] discuss how social media plays a crucial role in understanding public health. Paul et al. [11] go on to show that the current important health topics can be mined using social media data. Balani et al. [12] use social media to understand mental health issues using self-disclosed tweets, which is a methodology we also adopt in this paper. Choudhary et al. [13] use social media data to predict depression, indicating that social media interactions and posts serve as a good indicator for predicting mental health issues. McIver et al. [14] study prevalence of sleep issues using self disclosures on social media data. Zhang et al. [15] use Twitter self disclosures to understand recovery and relapse from alcohol use disorder. They use tweet text and peoples' social network to predict whether a user will recover from or relapse into alcohol use disorder. Cyberbullying is another problem that has garnered attention in the recent years on social media [16], [17], [18]. Twitter has also been used to study political events and related stance [19], [20], human trafficking [21].

These existing work helps us in identifying the important problems that could plague people during this time and will likely be exacerbated by the pandemic.

III. Data and Methods

In this section, we discuss our data collection methodology to investigate the socio-economic distress caused by COVID-19.

A. Data Collection

We collect data using the Twitter search API for a period of 3 months from March 14 to June 9. As our goal is to conduct a linguistic analysis to understand the impact of the pandemic on the socio-economic conditions of addiction, mental health, unemployment and immigration, we collect tweets related to these four categories, namely *1)* Addiction, *2)* Mental Health, *3)* Unemployment, and *4)* Immigration to quantitatively and qualitatively understand the context of these tweets. We collect only English language tweets and restrict our data collection to the United States as the authors have a better understanding of the government policies, pandemic-related actions and spread of the disease in the country during this time. Table I shows the number of tweets collected in each category. We observe that we were able to collect the maximum number of tweets related to unemployment and least with respect to addiction. This is understandable because COVID-19 necessitated stringent lockdowns and closures starting from March 2020, which resulted in significant number of job losses.

TABLE I: Number of Tweets by Category

Category	**Number of Tweets**
Addiction	3,324
Mental Health	5,754
Unemployment	32,100
Immigration	8,388

As a continuation of our previous work [1], we first collect data pertaining to a number of different COVID-19 hashtags from March 14 to May 4. We observe that there are several tweets related to addiction, mental health, unemployment and immigration in this dataset which we filter out using category-related keywords once the data is collected. For addiction, we use *'addict'* as the root keyword as it covers all its other forms. For mental health, we use *'mental health'* in its entirety as the keyword, while for unemployment, we use the words *'unemploy'* and *'jobless'*. For the immigration category, we use the root word *'immigra'* as it would cover all the other forms of the word immigrant.

We observe a decrease in the number of relevant tweets obtained using this general data collection approach over time. Figure 1 shows the number of tweets collected per day for the various categories. We observe from the figure that the number of collected tweets decreases in the beginning of May. Therefore, we adopt a more targeted data collection strategy from May 5 to June 9. The Twitter Search API has a provision for searching and collecting tweets using a combination of hashtags and keywords. We use #covid19 as a common hashtag along with keywords mentioned earlier for the different categories starting from May 5.

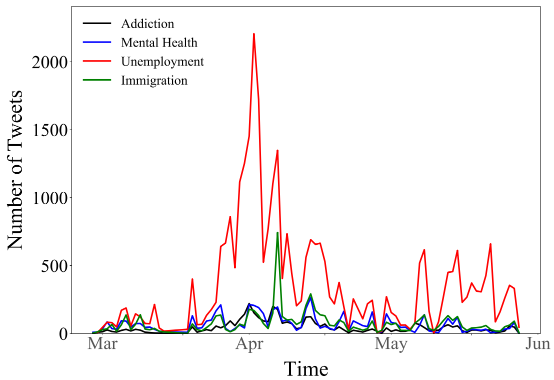


Fig. 1: Data Collection Timeline

We present some example tweets in Table II to illustrate the types of communications related to these categories occurring on Twitter during this period. We also observe that some of the tweets have political connotations with some users rallying for Donald Trump, the then US President, while others disagreeing with the administration's actions.

B. Gaps in Data Collection

Since we use the standard Twitter search API for data collection, the results are limited to just seven days, while more extensive records are available to premium and enterprise account holders. Second, the standard search API is rate limited. That is, there is a limit to the number of times an individual user or application may execute a specific action within a given time frame. Because our data collection utilizes application authentication and the 'GET search/tweets' method, we are limited to 450 search requests every fifteen minutes. We note that an individual request in this case can return multiple tweets, and it is therefore possible to stay within the rate limit and collect a sizable amount of data. The challenge lies in maximizing automated data collection without exceeding rate limits, as Twitter will block and may permanently ban users and applications that violate these limits. Though our data collection tool was never permanently banned, it was periodically blocked for exceeding rate limits. This resulted in occasional gaps in data collected via the standard search API. Third, as we use a keyword search method for obtaining tweets in the latter half of the study, it is likely that despite our best efforts we have inadvertently missed some related tweets because users have used different set of words to convey a similar meaning.

As is the case with most studies based on Twitter data, we also acknowledge the presence of bias in data collection [22]. The primary goal of our work is to interpret and present a comprehensive picture of peoples' opinion on deep societal issues such as unemployment, immigration, mental health, and addiction via users' self-expressed tweets as the pandemic upended their lives, so that we can collectively learn and do a better job in future. Therefore, despite some of these limitations, we are confident that the results presented here help in understanding the problems faced by the people during the pandemic.

IV. Data Analysis

In this section, we present observations and results based on our analysis of the tweets. We begin by investigating unigram (i.e., term) and bigram (i.e., co-occurrence) frequencies and then perform semantic role labeling and contextual analysis to identify the top action words and the associated nouns, respectively. We design a seeded topic modeling approach to automatically discern the main topics of discussion in each category. We conclude our analysis by determining the underlying sentiment of the tweets in the different categories. Our linguistic analysis identifies the general mood, concerns and sentiment of the public with regards to addiction, mental health, unemployment and immigration as the pandemic gripped the world and lays the foundation for more insightful analysis in the future.

A. Linguistic Word-Usage Analysis

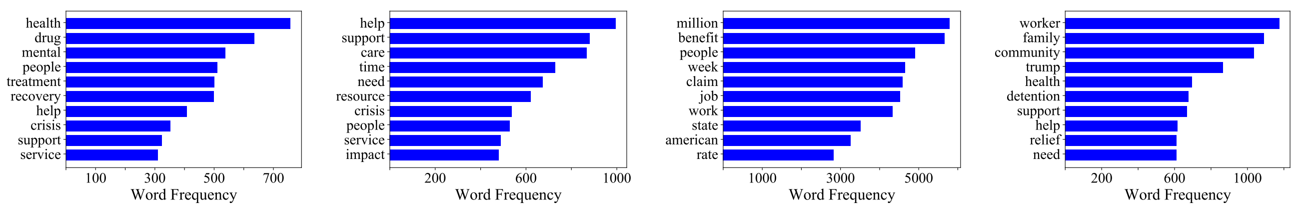
We begin our linguistic study by determining the most frequent words that occur in each category. As the first step, we perform data pre-processing to remove all trivial information from the data such as stop words, symbols, emojis, and web links. We create a common word list containing commonly occurring words across all categories and exclude these words from each category. For example, our common word list contains words such as COVID, pandemic, and corona. Figure 2 shows the frequently occurring words in each category. In the addiction category, we observe that *'drug'* is the second most frequent word, which indicates a possible occurrence of words signifying drug usage along with addiction. Other top words include *'recovery'* and *'treatment'*, which shows that people are concerned about the virus negatively impacting their recovery and treatment plans. Similarly, for mental health, we see words such as *'help'*,*'support'*,*'need'*, and *'care'*, which indicates that people are talking about asking for support and help. The unemployment top words resonate with the pandemic-induced job losses with millions of people trying to claim unemployment benefits (as seen from words such as *'million'*, *'claim'*, and *'benefit'*). Top words in the immigration category include *'family'*, *'community'*, *'detention'* and *'workers'*, signifying the different issues faced by immigrant workers and their families due to the pandemic. Word frequencies provide us a first-order approximation of the topics being discussed in the tweets and we use this to further guide our examination and analysis.

B. Word Collocation Analysis

As a next step in our lexical word analysis, we analyze co-occurrence of words (i.e., bigrams) to retrieve the contextual information surrounding the words. Before applying the algorithm to extract bigrams, we remove all adjacent spaces and stop words which are not meaningful. We perform stemming and lemmatization to limit the extended forms of words to the root word, so that these words can be analyzed

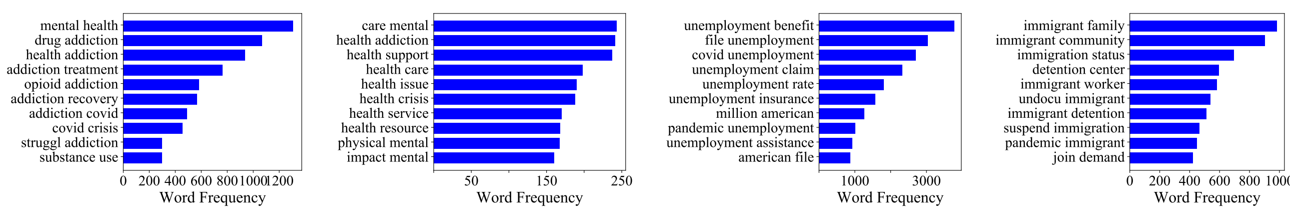
TABLE II: Example Tweets by Category

Category	Example Tweets
Addiction	We know that addiction is a symptom of disconnection, and the need for #socialdistancing can be especially difficult if you're struggling with alcohol or other drugs. You are not alone. You do not need to wait to reach out. #COVID19 #recover
Mental Health	This quarantine is seriously affecting the mental health of my whole family. #coronavirus
Unemployment	They're creating unemployment and purposely collapsing the economy to undo anything positive Donald Trump has done
Unemployment	Laid off? Stuck at home? Jobless? Now is the best time to take action on those business ideas you have always had! Make a difference and be the change! #ThursdayMotivation #ThursdayThoughts #entrepreneurs #coronavirus
Immigration	Just another example of how the Trump administration sees immigrants as a subpar species. Disgusting! #ShutItDown #COVID19



(a) Addiction (b) Mental Health (c) Unemployment (d) Immigration

Fig. 2: Word Frequencies



(a) Addiction (b) Mental Health (c) Unemployment (d) Immigration

Fig. 3: Bigrams

as a single item. We then calculate the overall frequency of each bigram and calculate the bigram association using the Pearson's Chi Squared Independence test, which evaluates if a pair of words have a higher probability of occurring together than by random chance. We present the top 10 bigrams in each category in Figure 3. We clearly see how bigrams capture more contextual information when compared to the unigrams. In the addiction category, we observe that bigrams such as *'drug addiction'*, *'substance use'*, and *'opioid addiction'* clearly articulate some of the main issues being faced by users in that group. Similarly, in the unemployment category, we see *'file unemployment'*, *'unemployment claim'*, and *'unemployment assistance'* emerging as the top bigrams that point to the high levels of unemployment and the issues that people are facing while obtaining unemployment benefits. In the immigration category, we observe bigrams such as *'immigration status*, *'detention center'*, *'undocumented immigrant'*, and *'suspend immigration'*, which highlights the plight of immigrants caused by measures such as tightening of borders, suspended immigration and travel that are adopted as a means to contain the rapidly spreading virus.

C. Semantic Role Labeling

We use an AllenNLP BERT-based semantic role labeling model to identify the action words in each category, which helps us automatically discern what people are trying to convey in their tweets. To obtain meaningful key action verbs in each category, we remove all the commonly occurring verbs across the categories using term frequency- inverse document frequency (TF-IDF) vectorization. Next, we compute the verb frequency of the existing words in each category to find the top verbs. Figure 4 shows the top verbs and their frequencies.

We observe that the addiction category contains words such as *'struggle'*, *'stop'*, *'deal'*, and *'lose'* that embodies the concerns and challenges faced by the people as the virus gripped society. We also observe words such as *'achieve'* and *'improve'* that come from tweets that are focused on spreading positivity to help those trying to cope with their addiction. The mental health category contains verbs such as *'manage'*, *'improve'*, and *'cope'* that are trying to motivate people who are affected, while other words such as *'affect'*, *'impact'*, and *'deal'* directly points to people's state of mental health. The unemployment category features verbs such as *'lose'*,*'need'*, *'work'* and *'lay(off)'*, which directly relates to the increased job losses and need to find new jobs. Words such as *'leave*, *'suspend'*, and *'lose'* are important verbs that feature in the immigration category and epitomize the concerns (e.g., leaving the country, losing their status) of immigrants during the pandemic.

D. Contextual Analysis of Action Words

Having identified the action words (i.e., verbs), we analyze the words associated with these verbs using dependency parsing to understand the context in which these verbs are used. Dependency parsing is a linguistic tool that parses sentences, isolates the key components, and represents them in the form of a tree. Typically in a dependency parse tree, the action words (i.e., verbs) occur in the root of the tree and subtrees, with the branches capturing the different contexts in which the action words are used in the tweet. We extract the sub-tree where a action word of interest is present and analyze the link associated with the nouns associated with the action words. We observe that 'nsubj' (Nominal Subject), 'pobj' (Object of a preposition), and 'dep' (The Dependent Object) are the link tags that contribute most to the action words. We use AllenNLP's dependency parser for the purpose of our study. The dependency parser is a neural network model using biaffine classifiers on top of a bidirectional LSTM [23]. Figure 5 gives an example of a dependency graph for each of the categories. The associated noun for the action word *suffer* is *people* as indicated in Figure 5b and the action word *file* is associated with *claim* as indicated in Figure 5c. Thus, understanding these noun-verb dependencies can help in understanding the context of different tweets better.

TABLE III: Addiction Action Words

Action words	**Linked Nouns**
continue	recovery, treatment, health, hand, people
reach	support, hope, care, friend, attempt
stop	coverage, control, blame, program, farce
struggle	addiction, media, health, depression, pandemic
deal	substance, crisis, recovery, people, anxiety

TABLE IV: Mental Health Action Words

Action words	**Linked Nouns**
need	help, assistance, support, staff, food
affect	community, people, wellness, family, unemployment
cope	stress, anxiety, health, distress, substance
understand	time, consumption, survey, study, life
deal	isolation, stress, homelessness, criticism, people

TABLE V: Unemployment Action Words

Action words	**Linked Nouns**
need	assistance, aid, help, word, file
find	work, job, information, people, combat
lay	people, job, work, file, insurance
receive	payment, money, news, stimulus, check
hit	rate, economy, job, record, industry

We retrieve the top 5 verb-noun combinations to understand the context surrounding the action words. This gives a better contextual understanding of the tweet as verbs and their associated nouns are the most important parts of speech that constitute a sentence. Tables III, IV, V, and VI show the top verb-noun combinations for the addiction, mental health, unemployment, and immigration categories, respectively. In the addiction category, we observe that the pandemic has compounded the challenges faced by individuals as indicated by the top verb-noun pairs such as *'continue recovery'*, *'struggle addiction'* and *'struggle pandemic'* making the list. We observe that a variety of issues are adversely affecting the mental health of individuals as is evidenced by the following verb-noun pairs—*'need food'*, *'affect unemployment'*, *'cope distress'*, and *'deal homelessness'*. We observe people discussing layoffs (e.g., *'lay job'*), needing help with filing claims (e.g., *'need file'*) and receiving stimulus (e.g., *'receive stimulus'*, *'receive check'*) in the unemployment category. Finally, Table VI outlines the problems faced by immigrants and the support they are getting from people across the US.

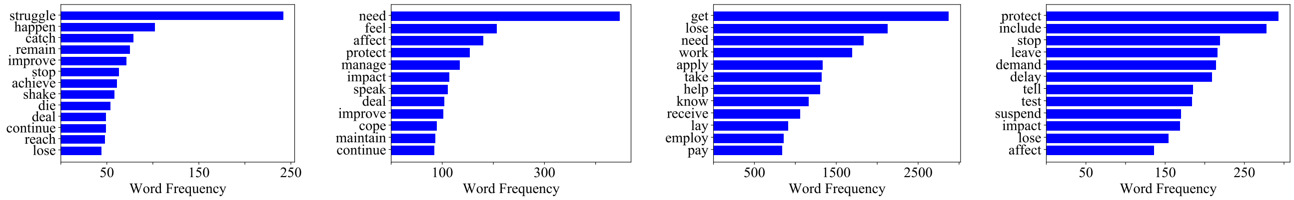
TABLE VI: Immigration Action Words

Action words	**Linked Nouns**
want	country, freedom, control, status, judge
suspend	order, news, plan, country, administration
impact	information, lesson, territory, reason, motto
hear	legislation, debate, office, gear, project
break	border, trial, jail, heart, nation

E. Seeded Topic Models

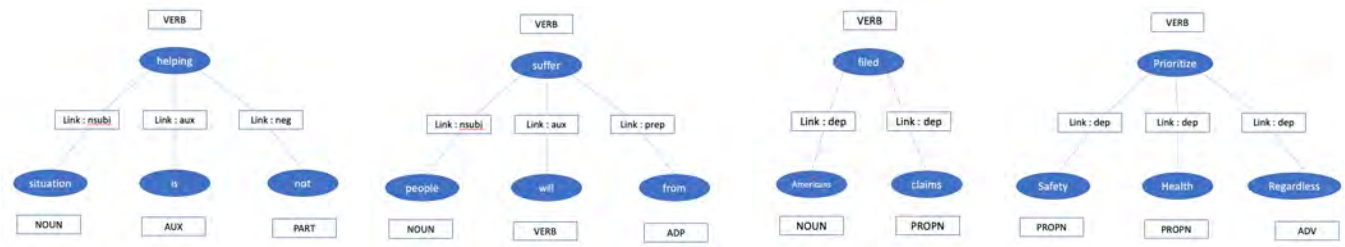
In this section, we adopt a seeded topic modeling approach, Seeded Latent Dirichlet Analysis (LDA) [24], to identify the main topics of discussion in each category and to automatically categorize tweets to a topic. Adopting a seeded LDA model enables us to design a scalable system that can help discern the main topics of discussion in each category, thereby complementing and strengthening the linguistic analysis conducted thus far. Seeded LDA, is a seeded variant of LDA, where the seed words are used to guide the topic discovery. The seed words are not exhaustive, but they provide direction to the topic discovery by influencing the document-topic and topic-word distributions.

We execute the Seeded LDA model for each category separately. In order to identify the most relevant topics of discussion, we leverage the top verbs and linked nouns in each category and use some of them as seed words. In addition to k seeded topics, we sometimes include an unseeded topic in our model to account for tweets that do not fall into the k topics. After experimenting with different values of k and manually evaluating the topics, we find the optimal k for each category that provides the best separation. We use α = 0.1 and β = 0.01 to give us sparse document-topic and topic-word distributions where fewer topics and words with high values emerge, so we can classify the tweets to the predominant category. We train the seeded LDA models for 2000 iterations. Across the categories, we find that the topics are related to the seed words for the most part and only a proportional number of tweets are assigned to the unseeded topics. Table VII, VIII, IX, X, show the seed words for the addiction, mental health, unemployment, and immigration categories, respectively. For each of these categories, we also discuss the words that occur in the top words for each topic as identified by the seeded



(a) Addiction (b) Mental Health (c) Unemployment (d) Immigration

Fig. 4: Top Verbs



(a) Addiction (b) Mental Health (c) Unemployment (d) Immigration

Fig. 5: Dependency Graph

LDA model in addition to the seed words in Tables XI, XII, XIII, and XIV.

TABLE VII: Addiction Seed list

Topic	Seed List	# Tweets
Mental State	depression, anxiety, stress, mental, suicide	1774
Struggles	struggl, recovery, support, relapse	1703
Misc	unseeded	716

TABLE VIII: Mental Health Seed List

Topic	Seed List	# Tweets
Assistance	support, help	1817
Addiction	addiction, alcohol, drug	940
Quarantine	isolation, distance, loneliness	1664
Misc	unseeded	1451

TABLE IX: Unemployment Seed List

Topic	Seed List	# Tweets
Claims	unemployed, job, insurance, lose, file	9289
Aid	aid, assistance, relief, stimulus	8561
Economy	economy, business, government	9309
Misc	unseeded	11207

TABLE X: Immigration Seed List

Topic	Seed List	# Tweets
Illegal	detention, order, ban, undocumented	4448
Aid	support, relief	4276

In the addiction category, we identify two seeded topics, specifically **'mental state'** and **'struggles'** along with a miscellaneous unseeded topic that accounts for all the remaining documents (i.e., tweets). Our first topic is **'mental state'**, and the seed words (e.g., *'depression'*, *'anxiety'*, *'stress'*, and *'suicide'*) align with the topic and demonstrate that people are talking about serious mental health issues such as depression, anxiety, and stress together with addiction. Apart from the seed words, we also observe words such as *'abuse'* and *'overdose'* in this topic (Table XI). In the second topic, **'struggles'**, we come across words such as *'offer'*, *'cope'*, and *'provide'* in the top words, which are related to our seed words *'support'*, *'recover'*, and *'relapse'*. In the unseeded topic, we notice words such as *'alcoholism'*, *'worse'* and *'medicine'*, which indicates the worsening condition of addictions.

In the mental health category, the first topic is **'assistance'** and key words pertaining to the topic are *'support'*, *'wellness'* and *'program'*. This topic captures tweets expressing need for emotional support and care along with tweets that offer and provide help. The second topic is **'addiction'** and we see words such as *'drug'*, *'unemployment'*, *'suicide'*, *'violence'* and *'abuse'*. These words refer to people discussing addiction-related issues involving alcohol and drugs, which consequently might impact their mental health leading to drug abuse, suicide and unemployment. It is important to note that our analysis reveals that while in the addiction category, mental health is a prominent sub-topic, in the mental health category, addiction related issues are a predominant sub-topic, suggesting a close relationship between these two issues of addiction and mental health. The third topic in this category is **'quarantine'** and words such as *'isolation'* and *'loneliness'* in this topic indicates the prevalence of mental health issues elevated by quarantine and lockdown measures. On the other hand, the occurrence of words such as *'exercise'* and *'physical'* in the context of mental health are suggestive of people participating in exercising and outdoor activities, which have been proven to improve mental health.

TABLE XI: Addiction Word List

Topic	Top Words
Mental State	treatment, abuse, urge, overdose, risk
Struggles	difficult, cope, achieve, provide, offer

TABLE XII: Mental Health Word List

Topic	Top Words
Assistance	manage, access, resource, wellness, program
Addiction	suicide, abuse, unemployment, substance, violence
Quarantine	anxiety, physical, depression, outside, virtual, fear

TABLE XIII: Unemployment Word list

Topic	Top Words
Claims	labor, lost, risk, claim, laid, fraud, administration
Aid	payment, receive, support, fund, benefit, federal
Economy	rate, market, impact, recession, stock, wage, historic

TABLE XIV: Immigration Word List

Topic	Top Words
Illegal	immigrant, delay, suspend, deportation, illegal
Aid	essential, fund, stimulus, assistance, impact, federal

The first topic in the unemployment category centers around unemployment **'claims'**. Here, we observe words such as *'unemployed'*, *'job'*, *'insurance'*, *'lose'* and *'file'*. The words *'claim'* and *'file'* in this topic refer to filing for unemployment benefits. The second topic of discussion in this category involves **'aid'**. Words such as *'federal*, *'stimulus'*, *'payment'*, *'support'*, *'fund'* and *'benefit'* capture the representative words in the topic. The presence of these words in the tweets convey that people are relying on aid or federal stimulus to get through this difficult phase of unemployment complicated by the pandemic. The third topic here captures the **'economy'**. The words representative of this topic are *'recession'*, *'wage'* and *'historic'*.

In the immigration category, the first topic captures discussions around **'illegal'** immigration. The words representative to this topic are *'suspend*, *'illegal'*, *'undocumented'* and *'ban'*. These words correspond to discussions around illegal immigrants, suspending immigrant workers, and the risks they are facing. The second topic is **'aid'**. The tweets in this topic raise and discuss issues such as many immigrant workers being *'essential'*, and the need for *'stimulus'* that also includes immigrant workers. Also, since these two topics capture most of the tweets in this category, we do not include an unseeded topic for this category in the seeded LDA model.

Using seeded LDA, our goal is to show the fine-grained connections among the different topics and the categories that exist in this data. Seeded LDA essentially gives us a better understanding of what people are discussing in the various categories. We also see similar topics such as '*aid*' and '*support*' repeating across the categories, which conveys that people are in need and looking for help and assistance.

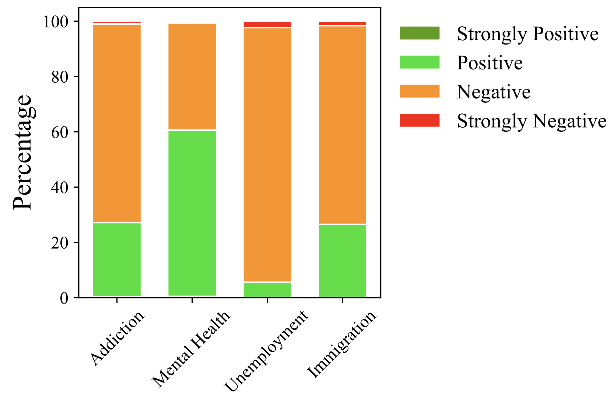


Fig. 6: Multi-class Classification Sentiment Graph

F. Sentiment Analysis

To understand the sentiment for the different categories, we perform a comparative sentiment analysis. We use a pre-trained sentiment analysis model [25], which has 95.11% accuracy on the Stanford SST test dataset and apply it to our dataset. Our sentiment classification model, roBERTa base model, classifies the data into five sentiment categories: strongly positive, positive, neutral, negative, and strongly negative. As tweets in the neutral category are generally statements with no positive/negative sentiment words, we exclude the tweets in the neutral category and scale the tweets in the other four categories to 100%.

Figure 6 shows the overall sentiment in the different categories. We observe that the overall sentiment in the addiction, unemployment, and immigration categories is primarily negative. The highest percentage of negative and strongly negative sentiment tweets is seen in the unemployment category. This is understandable because tweets in the category are primarily from people who have lost their jobs or fear losing their jobs due to the pandemic. The negative sentiment in the immigration category can be attributed to border closures and the fear of deportation as the United States tried to control the spread of the disease. While mental health is an issue that has plagued many people during the pandemic due to the increased stress levels, we observe that people took to social media to discuss several pertinent issues in a more positive light when compared to the other issues. We observe that overall sentiment in the mental health category is positive as majority of the tweets are related to taking good care of one's mental well-being during lockdown and quarantine.

V. Discussion and Concluding Remarks

In this paper, we studied Twitter communications in the United States during the first few months of the COVID-19 outbreak to understand the socio-economic disruption caused by the pandemic. With the pandemic resulting in lockdowns, closures, and social isolation, we investigated its impact on socio-economic conditions of addiction, mental health, unemployment, and immigration. Our fine-grained linguistic analysis revealed the distress of people as self-expressed by them on social media. By examining unigram and bigram frequencies, action words and their associated context, we unearth the problems that people experienced during these unprecedented times. For example, words such as *'stimulus'*, *'insurance'* and phrases such as *'file unemployment'* and *'receive stimulus'* in the unemployment category demonstrate the hardships experienced by the people as the pandemic upended their lives. Additionally, our seeded topic modeling approach identified the key topics of discussion in each category (e.g., the growing problem of addiction and mental health exacerbated by the pandemic, discussions around aid/help in multiple categories). Finally, by conducting sentiment analysis of the tweets, we observed that the sentiment is largely negative with respect to all these different categories, which once again points to peoples' struggles. As part of our future work, we are currently collecting data to understand the the acceptance/reluctance to the COVID-19 vaccines and issues related to the equity of the vaccine distribution.

References

[1] S. G. Shanthakumar, A. Seetharam, and A. Ramesh, "Analyzing societal impact of covid-19: A study during the early days of the pandemic," in *International conference on Social Computing and Networking*, 2020.

[2] W. Ahmed, J. Vidal-Alaball, J. Downing, and F. L. Seguí, "Covid-19 and the 5g conspiracy theory: social network analysis of twitter data," *Journal of Medical Internet Research*, vol. 22, no. 5, p. e19458, 2020.

[3] E. Ferrara, "What types of covid-19 conspiracies are populated by twitter bots?" *First Monday*, May 2020. [Online]. Available: http://dx.doi.org/10.5210/fm.v25i6.10633

[4] R. Kouzy, J. Abi Jaoude, A. Kraitem, M. B. El Alam, B. Karam, E. Adib, J. Zarka, C. Traboulsi, E. W. Akl, and K. Baddour, "Coronavirus goes viral: quantifying the covid-19 misinformation epidemic on twitter," *Cureus*, vol. 12, no. 3, 2020.

[5] H. W. Park, S. Park, and M. Chong, "Conversations and medical news frames on twitter: Infodemiological study on covid-19 in south korea," *Journal of Medical Internet Research*, vol. 22, no. 5, p. e18897, 2020.

[6] H. Budhwani and R. Sun, "Creating covid-19 stigma by referencing the novel coronavirus as the ?chinese virus? on twitter: Quantitative analysis of social media data," *Journal of Medical Internet Research*, vol. 22, no. 5, p. e19301, 2020.

[7] "Twitter discussions and emotions about the covid-19 pandemic: Machine learning approach."

[8] M. Pérez-Pérez, G. Pérez-Rodríguez, F. Fdez-Riverola, and A. Lourenço, "Using twitter to understand the human bowel disease community: Exploratory analysis of key topics," *Journal of medical Internet research (JMIR)*, vol. 21, no. 8, p. e12610, 2019.

[9] M. Dredze, "How social media will change public health," *Journal of IEEE Intelligent Systems*, vol. 27, pp. 81–84, 2012.

[10] M. J. Paul and M. Dredze, "You are what you tweet: Analyzing twitter for public health." in *Proceedings of the International Conference on Web and Social Media (ICWSM)*, 2011.

[11] ——, "Discovering health topics in social media using topic models," *Journal of PLOS ONE*, vol. 9, p. e103408, 2014.

[12] S. Balani and M. De Choudhury, "Detecting and characterizing mental health related self-disclosure in social media," in *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems*, 2015.

[13] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media." in *Proceedings of the International Conference on Web and Social Media (ICWSM)*, 2013.

[14] D. J. McIver, J. B. Hawkins, R. Chunara, A. K. Chatterjee, A. Bhandari, T. P. Fitzgerald, S. H. Jain, and J. S. Brownstein, "Characterizing sleep issues using twitter," *Journal of Medical Internet Research*, vol. 17, 2015.

[15] Y. Zhang, A. Ramesh, J. Golbeck, D. Sridhar, and L. Getoor, "A structured approach to understanding recovery and relapse in aa," in *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2018, pp. 1205–1214.

[16] S. Tomkins, L. Getoor, Y. Chen, and Y. Zhang, "A socio-linguistic model for cyberbullying detection," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2018, pp. 53–60.

[17] E. Raisi and B. Huang, "Weakly supervised cyberbullying detection using co-trained ensembles of embedding models," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2018.

[18] Y. Zhang and A. Ramesh, "Fine-grained analysis of cyberbullying using weakly-supervised topic models," in *International Conference on Data Science and Advanced Analytics (DSAA)*, 2018.

[19] K. Johnson and D. Goldwasser, "All i know about politics is what i read in twitter: Weakly supervised models for extracting politicians' stances from twitter," in *International Conference on Computational Linguistics (COLING)*, 2016.

[20] H. Le, G. Boynton, Y. Mejova, Z. Shafiq, and P. Srinivasan, "Bumps and bruises: Mining presidential campaign announcements on twitter," in *Proceedings of the Conference on Hypertext and Social Media*, 2017.

[21] S. Tomkins, G. Farnadi, B. Amanatullah, L. Getoor, and S. Minton, "The impact of environmental stressors on human trafficking," in *2018 IEEE International Conference on Data Mining (ICDM)*, 2018.

[22] F. Morstatter and H. Liu, "Discovering, assessing, and mitigating data bias in social media," *Journal of Online Social Networks and Media*, vol. 1, pp. 1–13, 2017.

[23] T. Dozat and C. D. Manning, "Deep biaffine attention for neural dependency parsing," in *International Conference on Learning Representations*, 2017.

[24] J. Jagarlamudi, H. Daumé III, and R. Udupa, "Incorporating lexical priors into topic models," in *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 2012, pp. 204–213.

[25] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.