

Article

Studying Unemployment Effects on Mental Health: Social Media versus the Traditional Approach

Samara Ahmed ¹, Adil E. Rajput ^{2,*}, Akila Sarirete ², Asma Aljaberi ³, Ohoud Alghanem ³ and Abrar Alsheraigi ³

¹ Psychiatry Division, College of Medicine, King Abdulaziz University, Jeddah 22252, Saudi Arabia; samraa2018@gmail.com

² Computer Science Department, College of Engineering, Effat University, Jeddah 22332, Saudi Arabia; asarirete@effatuniversity.edu.sa

³ Information Systems Department, College of Engineering, Effat University, Jeddah 22332, Saudi Arabia; asaljaberi@effat.edu.sa (A.A.); ohalghanem@effat.edu.sa (O.A.); aalsheraigi@effat.edu.sa (A.A.)

* Correspondence: aialahi@effatuniversity.edu.sa; Tel.: +966-507840247

Received: 14 August 2020; Accepted: 25 September 2020; Published: 2 October 2020



Abstract: Social media, traditionally reserved for social exchanges on the Internet, has been increasingly used by researchers to gain insight into different facets of human life. Unemployment is an area that has gained attention by researchers in various fields. Medical practitioners especially in the area of mental health have traditionally monitored the effects of involuntary unemployment with great interest. The question we want to address is as follows: while many researchers have been using data from social media and microblogging sites in the past few years, do they provide results consistent with traditional research? Furthermore, if the data are indeed consistent, are they detailed enough to deduce possible reasons and remedies? We believe that having a concise answer to these questions is imperative for a sustainable mechanism for medical practitioners and researchers to gather and analyze data. The stigma of mental health prevents a good portion of society from seeking help, but the anonymity provided by the Internet could shatter such barriers, thus allowing people affected by conditions such as mental health and unemployment to express themselves freely. In this work, we compare the feedback gathered from social media using crowdsourcing techniques to results obtained prior to the advent of social media and microblogging. We find that the results are consistent in terms of (1) financial strain being the biggest stressor and concern, (2) the onslaught of depression being typical and (3) possible interventions, including reemployment and support from friends and family, playing a crucial role in minimizing the effects of involuntary unemployment. Lastly, we could not find enough evidence to study effects on physical health and somatization in this work.

Keywords: social media; unemployment; crowdsourcing; natural language processing; mental health

1. Introduction

1.1. Motivation and Background

Unemployment has been a subject of interest among a wide community of researchers spanning various disciplines. Specifically, healthcare providers have long been studying effects of involuntary unemployment in societies both on an individual and aggregate level [1,2]. The studies have been mostly confined to a particular geographical area, as traditionally it has been difficult to do an aggregate study spanning various geographical levels. Furthermore, the literature suggests that various societies and countries have a different intervention system in place (unemployment benefits,

retraining opportunities etc.)—a fact that further complicates comparing samples across various regions and societies.

Interestingly, researchers have shown (in separate studies) that both unemployment and mental health has an associated stigma and comes with a feeling of shame for the affected person. Work done by [3,4] shows how mental health problems bring a feeling of shame to persons suffering from it while the authors in [5–7] discuss how financial hardship and unemployment is accompanied by a feeling of shame (especially for men traditionally). The change of working environment and increase in unemployment in the new economy has brought forth various challenges [8]. Hence, there is a need for mental healthcare practitioners to find a new paradigm for the society to sustain under such environment.

Social computing has emerged as a discipline where users employ various computing devices for social use rather than increasing productivity. The work done in [9] points out the change in social interaction among people in the past few years. The social media platforms offer a plethora of data in forms of free text that embodies opinions of users across the globe. Thorne, in [10] provides a comprehensive study where he argues how the Internet has helped in learning different languages. What is also interesting is the fact that the author in one case study observes that the difference in use of the communication tools on the Internet is similar when engaging in non-internet communication. Dunkels in his work [11] did a study on children's perceived dangers (Swedish 6th graders between the age of 11 and 13) of the Internet and the fact that children of various ages not only were aware of the dark side of online communications but also had various counter strategies very different than those of adults and their parents. Savin et al. [12] use the shortage of child and adolescent psychiatrists across the globe as motivation and make a case of telepsychiatry using video conferencing. The authors specifically study whether the difference in culture could impede in providing effective treatment to the patients. The study concludes that many impediments can be overcome by paying attention to certain details such as studying nonverbal communication clues, historical aspects etc. The behavior was observed across various cultures during COVID-19 when doctors across the globe used video conferencing as a means to connect to their patients.

Coupling the information gathered from social media where various communities exist to exchange ideas and express one's feelings with the power of Big Data, the question that comes to mind is whether the emergence of various social media tools helped in the convergence of uniform expression of symptoms and the effects of a global human condition such as unemployment. The traditional literature on unemployment pays great attention to various intervention factors such as community support in minimizing the negative effects of unemployment. The existence of social media tools and blogs are known to conflate exchanges across the globe, but do they truly reflect the actual thought process of people that are affected? Work done by [13] showed that the profile created by Facebook was more accurate than the portrait given by actual friends of a person.

This leads to another question: how does one gather and group data for a particular topic? The concept of crowdsourcing (originally meant for outsourcing tasks to humans regardless of their physical location) has been employed increasingly to group relevant data together. Paniagua and Korzynski [14] show how the concept of crowdsourcing is part and parcel of the social media platform. Thus, the Twitter platform allowed, in an emergency situation, the gathering of feedback from volunteers in the affected region and helped the authorities to respond appropriately [15]. This was a case where the volunteers were fully cognizant that their input on social media is being used to form an action plan and is considered an example of active outsourcing. On the other hand, passive crowdsourcing gathers users' thoughts without the users recognizing that they are contributing. The concept of hashtag on Twitter, where various users contribute to a given subject, falls under the concept of passive crowdsourcing. Many social media providers provide mechanisms that allow researchers to gather data from such site. The researchers employ techniques developed in the field of Natural Language Processing (NLP) to segment and cull together relevant text gaining an insight in

both the context and the relevant meaning. The NLP techniques depend on an existing corpus that can represent a language [16].

Demzsky et al., in [17], employ NLP techniques to uncover linguistic dimensions of political polarization. The authors use the social media data to confirm earlier literature relating to conceptualization of race in US. While the general corpus represent a particular language/dialect in general, various domains are represented by a corpus more specific to the problem at hand. Rajput and Ahmad in [18] make a case for a corpus to assist mental health professionals in detecting depression among users provided some group of people. This is crucial to mental health realm as studies such as [19] have shown that a significant correlation exists between depression and suicidal behavior and depression is an ailment that affects the global community uniformly. The researchers make use of the Twitter hashtag #depression and conclude that keywords gleaned mimic the language of depression patients. Such a corpus can serve to segment random text to predict with certain assurance the prevalence of depression symptoms within the thoughts described by a particular user.

1.2. Problem Description

Our current work aims at scavenging data from the social media using NLP techniques and crowdsourcing concept to gather and analyze data specific to unemployment. The work at hand had started before the COVID-19 crisis but the data collection coincided during the early part of the pandemic. The motivation for our work stems from the question: can social media platforms provide sufficient data that conform to the results obtained using the traditional methods. Specifically, we will try to categorize the results in terms of possible reasons, consequences and intervention techniques. We will further narrow down the consequences to see whether the users suffering from unemployment suffer any mental health symptoms.

We utilized the hashtag #unemployment and gathered data for one month to arrive at our results. The discussion was in English but contained several Out of Vocabulary (OOV) words that we ignored. We look at the keywords both individually and with other words. Given the importance of involuntary unemployment on a global scale, we try to address the following questions in this work:

1. RQ1: Other than the financial strain, does involuntary unemployment affect the users globally the same way when it comes to mental health?
2. RQ2: Can researchers use the data from social media as a basis for analysis compared to traditional analysis approach?
3. RQ3: Does the data scavenged from social media provide basis for both the consequences and intervention techniques when it comes to unemployment?

2. Literature Review

2.1. Traditional Measures for Unemployment

In [20], the authors discuss the long-term effects of unemployment on youth's behavior employing a longitudinal study. Discussing from a purely behavioral and economic perspectives, the authors argue that (1) the youth unemployment forces the youth to seek skills improving behavior by engaging in more training and (2) even a six-month unemployment negatively affects the income over a period of at least ten years. The authors argue that the unemployment propels the youth to accept opportunities offering less than their worth in fear of possible unemployment in the future. Arulamaplam in [21] confirmed these findings and concludes that, in Britain, the unemployment leaves a permanent financial scar on the individual and individuals earned 6% less on reentry while they earn 14% less after three years. Kessler et al., in [22], did a community survey and focused on the selection bias in earlier studies. After minimizing the selection bias issues, the authors concluded that, similar to prior studies, unemployment had a clinically significant impact on the health of unemployed individuals. They further concluded that (1) financial strain caused the biggest impact on health as financial strain's absence halved the negative health effects and (2) the unemployment compounded the effects of

otherwise unrelated life events. The clinically significant health effects included both physical and mental health strains such as depression and anxiety. Furthermore, the authors also mention the effects of somatization that are hard to measure. The authors also bring forth the effects of mediating factors that include reemployment, family support etc. Mastekaasa in [23] took an alternative view in literature where he argues, based on a study in Norway, that mentally healthy people are less likely to get laid off or have higher chance of reemployment fairly quickly while mentally ill people are at a higher risk for getting laid off. Authors in [24] take an opposite approach and prove that on an aggregate level there is a high correlation between long-term employment and suicide rate, while, on an individual level, depression and substance abuse is a common consequence of long-term unemployment. More recently, work done by Pohlan [25] confirms that unemployment has negative effects on different aspects of an individual's life, including social integration, life satisfaction, access to economic resources and more importantly an individual's mental health as it leads to social exclusion and eventually isolation from society. The authors also showed that having a partner and being highly educated reduces the negative effects of job loss. The work done by Voßemer et al. [26] takes a step further and limits the effects of unemployment to mental health and well-being but not on physical health. The work done in [27] concludes that, while the importance of education in the modern world is paramount, highly educated individuals have difficulty finding jobs appropriate to their level and end up taking employment that is lower than their academic qualifications, impacting their mental health negatively. The results were replicated by [28] in India. Using data from the German Socio-Economic Panel Study (SOEP) from 2002 through 2010 highlighted the negative effects on unemployment spouses' mental health and the fact that unemployment had severe consequences on both the unemployed and their spouse [29]. The authors in [30] looked at the correlation of sensory processing patterns and their relationship with underlying medical conditions. They concluded, among other things, that there is a significant correlation between unemployment and depression.

2.2. Social Media and Unemployment

Kunze and Suppa in [31] concluded that unemployment negatively affects social participation for public activities and exercises. However, they also reported that social media help unemployed people keep their relationships. The social network impacts on individual inclusion and exclusion as the unemployed people use social media to grow their social networks, and the chance to establish new contacts. Thus, social networks differentiate between unemployed and employed through persisting online [32]. The authors of [28] presented a contradicting view in terms of Information Technology's (IT) contribution towards unemployment, where certain studies supported that IT contributed towards unemployment while others view IT positively in terms of helping to find various avenues for employment including expanding social circles. The work done in [33,34], p. 2 explore the outsourcing of jobs in general and IT specifically to see the effects on unemployment. Proserpio et al., in [35], did a very thorough study of 230,000 US users that had either lost their jobs or gained a new one over five years from the year 2010 to 2015. The authors argued that psychological well-being elements can be used as leading indicators showing the economic indices weeks in advance with greater accuracy. Our research takes a similar approach but our goal is to see whether the social media mimics the results found by traditional methods. Suphan et al. [32] focused on the effective role of social media in reducing unemployment where a survey of 809 Facebook users showed that the unemployed people found it easier to use their virtual contacts as compared to the population of the rural regions. The community of the urban areas is at higher exposure to drop out in the previous social networks that can lead to poor mental condition due to the problem of unemployment. Mincer in [36] goes a step further in contending that minimum wage earners suffer similar health effects as those who suffer from involuntary unemployment.

The research done by [37] used social media content containing the news articles, blogs, and tweets written in the Korean language, extracted the social moods and predict the unemployment.

The study conducted by [35] in May 2015 found the relationship between psychological wellbeing and unemployment, by analyzing Twitter posts from United States users who either lost a job or gained

a new job. Our study would build upon this and some of our prior research and see whether we can pinpoint some intervention techniques. Furthermore, our research does not distinguish between the geographical location but rather see whether we can locate the areas most affected by unemployment. The work done by [38] focuses on people in the United States and contends that unemployed use social media more at night while employed people use it more during the day. Authors in [39,40] focus on predicting unemployment rated by employing NLP techniques.

2.3. Crowdsourcing

Crowdsourcing has gained popularity helping entities gather services, ideas, or content from a large group of people mostly using electronic channels. It is interesting to note that prior to social media, there were efforts done to cull together heterogenous data from various sources [41]. The advent of Peer-to-Peer networks also were a step forward to conflate data in an unstructured form [42]. Wazny, in [43], offers a thorough review of crowdsourcing, including taxonomy, prevalent research and various regulatory and ethical aspects. The author points out the potential of using crowdsourcing in a medical field as it has the potential to gather a vast amount of relevant data. The work done in [44] reviews crowdsourcing from various sources and presents an integrated definition. Ranard et al., in [45], present a systematic review of use of crowdsourcing [46] in the field of medicine. Similar work is also presented in [47]. Authors in [48], discussed the use of the Internet as a medium for crowdsourcing and discussed four significant challenges that included how to recruit, measure their abilities, maintain quality of the work and more importantly how to integrate the work performed. The work done in [49,50] uses the same concept to predict the socioeconomic status (SES), while the work done in [51] addresses the problem of female bullying at work. The work done in [18] uses a similar technique to detect depression but the work is more focused on building corpus. The work done in [52] describes how voluminous data are being produced in medical sciences and how intelligence can be gathered to resolve various ailments.

2.4. Big Data/Social Media and Mental Health

Murdoch and Detsky [53] introduced the need for applying the Big Bata techniques to medical field to gather better insights. Chen and Wojcik described a framework on applying Big Data in the field of psychology and mental health [54]. The authors focus on the four steps necessary for such endeavors, namely planning, acquisition, analysis and analytics, and also provide three tutorials for the users. The field of psychiatry saw many initiatives recently spurred by the interest in Big Data. These include developing suicide risk algorithms, risk of dementia, substance abuse disorders, prescribing psychotropic drugs and studying cognitive impairment. Monteith et al. summarizes the above and describes the ramifications Big Data is having in the field of psychiatry in [55], while the work done in [56] provides an overview of work done in medical sciences along with various techniques that are employed. Dechoudhry et al. has laid the foundation for work in applying such techniques to the social media platform [57,58]. Specifically, they have demonstrated predicting depression in Twitter users given a set of users who have indicated prevalence of depression in their lives. The work, however, is based on users who have indicated the prevalence of depression and made their tweets available.

3. Experimental Setup

As is the case in such research, we only collected the data from public sources (specifically Twitter) to ensure that we address the privacy concerns of using such data [59,60]. Furthermore, we do not publish any user handle on Twitter but make sure that we eliminate duplicate tweets to ensure that we are working from a clean set of data.

3.1. Preprocessing and Processing Data

We followed the following process for preprocessing and processing of data. Snippets of Python code is provided for the steps (The code is available to interested readers):

1. Collected over 25,000 tweets under the hashtag #unemployment.
2. We use the Natural Language Toolkit (NLTK) to parse the texts and get rid of the stop-words (recurring words such as articles that need to be filtered out) (Box 1).

Box 1. API call to NLTK tf-idf.

```
import csv
import time
import logging
from collections import defaultdict
import pandas as pd
import datetime
from datetime import tzinfo
import pytz
import requests
from nltk.corpus import stopwords
from numpy import argsort
from pytz import timezone
from dateutil import parser
from twitter import *
from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from numpy import argsort, squeeze
import numpy as np
LOCAL_TIMEZONE = pytz.timezone("US/Central")
TWITTER_TIMEZONE = timezone("UTC")
START_DATE_FOR_SEARCH = LOCAL_TIMEZONE.localize(datetime.datetime(2020, 4, 7, 14, 0))
END_DATE_FOR_SEARCH = LOCAL_TIMEZONE.localize(datetime.datetime.now())
# hashtag to search
HASHTAG = "#unemployment"
#from stopwords import STOP_WORDS
STOP_WORDS = ['https', 'rt', 'for']
```

3. We use the tf-idf algorithm [61] to generate the keywords (Box 2).

Box 2. Setting up the n-gram range.

```
tfidf_vectorizer = TfidfVectorizer(
    max_df=0.90,
    min_df=2,
    max_features=500,
    ngram_range=(1, 3),
    stop_words=[] #STOP_WORDS
)
```

4. We used the n-gram model for $n = 1, 2$ and 3 . This helped us get the top keywords (1-gram), 2 adjacent words (2-gram/bigram) and 3 adjacent words (3-gram/trigram). We use the NLTK built-in functionality (Box 3).

Box 3. Getting top key words.

```

def get_top_n_words(corpus, n=None):
    """
    List the top n words in a vocabulary according to occurrence in a text corpus.
    get_top_n_words(["I love Python", "Python is a language programming", "Hello world", "I love the world"]) ->
    [('python', 2),
     ('world', 2),
     ('love', 2),
     ('hello', 1),
     ('is', 1),
     ('programming', 1),
     ('the', 1),
     ('language', 1)]
    """
    vec = CountVectorizer(stop_words=stopwords.words('english') + STOP_WORDS, ngram_range=(1,3)).fit(corpus)
    bag_of_words = vec.transform(corpus)
    sum_words = bag_of_words.sum(axis=0)
    words_freq = [(word, sum_words[0, idx]) for word, idx in vec.vocabulary_.items()]
    words_freq = sorted(words_freq, key=lambda x: x[1], reverse=True)
    return words_freq[:n]

```

5. Once we have finalized the preprocessing part, we used the sklearn library to tokenize and vectorize the tweets.
6. For the sake of our work, we treated the entire set of tweets as one corpus (Box 4).

Box 4. Passing Twitter corpus to tf-idf.

```

def tf_idf(documents):
    """
    :param documents: Documents to extract TFxIDF
    :param top_n: top number of features to return
    :return: top TFxIDF features
    """
    df = pd.DataFrame(documents['documents'])
    X = tfidf_vectorizer.fit_transform(df['text'])
    indices = np.argsort(tfidf_vectorizer.idf_)[-1]
    features = tfidf_vectorizer.get_feature_names()
    features = [features[i] for i in indices[::-1]]
    print('TOP TFxIDF features')
    print(features)
    features_by_gram = defaultdict(list)
    for f, w in zip(tfidf_vectorizer.get_feature_names(), tfidf_vectorizer.idf_):
        features_by_gram[len(f.split(' '))].append((f, w))
    for gram, features in features_by_gram.items():
        top_features = sorted(features, key=lambda x: x[1], reverse=True)[-1]
        top_features = [f[0] for f in top_features]
        print('{}-gram top:'.format(gram), top_features)
    return features
    # df = pd.DataFrame(documents['documents'])
    # tfidf_model = tfidf_vectorizer.fit(df['text'])
    # tfidf = tfidf_model.transform(df['text'])
    # feature_names = tfidf_model.get_feature_names()
    # for i, _ in enumerate(tfidf):
    #     documents['documents'][i]['top5_tfidf'] = top_tfidf_features(squeeze(_.toarray()), feature_names, 5)
def top_tfidf_features(row, features, top_n = 25):
    topn_ids = argsort(row)[-1][:top_n]
    top_feats = [(features[i], row[i]) for i in topn_ids]
    df = pd.DataFrame(top_feats)
    df.columns = ['feature', 'tfidf']
    return df

```

7. In addition to collecting the n-gram keywords, we also collected all the hashtags that are mentioned in the tweets and the number of times they were used.

8. Implemented the above on a standard Dell running Ubuntu Linux and Python3 program with 16G RAM.

The Application Programming Interfaces (APIs) Used

For this work, we used the python programming to gather and analyze the data. We used the following open-source APIs available for python programming language:

1. Twitter API: This requires registering with Twitter and creating a twitter development account. The Twitter library can be installed for Python that provides all the requisite APIs.
2. Pandas: This is an open-source python library which allows data cleaning, preparation and fast analysis. The data can be easily imported into Excel.
3. NLTK: This is one of the most powerful NLP libraries that provides the basic tools such as tokenization, stemming, lemmatization, etc. Interested readers can refer to [62] for pertinent details.
4. Sklearn: This library helps in Big Data analysis such as classification, regression, clustering, etc.

4. Results and Discussion

Based on the results, we deduce the following:

1. The results obtained from the social media are indeed consistent with those of traditional research when it comes to unemployment. It is worth noting that while the social media platform (Twitter in our case) does not provide mechanism to segregate users based on factors such as gender, age, etc., the results we got were still consistent with the traditional research. (RQ1)
2. The tweets did provide enough data to deduce that involuntary unemployment in fact does have a universal effect on mental health of persons affected. (RQ2)
3. The data from social media does provide a basis for both the consequences and intervention techniques when it comes to unemployment.

The following subsection discusses this in more detail.

4.1. 1-Gram

As mentioned in the previous section, we looked at 1-gram, bigram and trigram keywords that we gleaned from the tweets. While the research for this work commenced prior to COVID-19 crisis, the tweets we gathered spanned the month of April and hence the results reflected this phenomenon. Below are the results of the top 20 terms using the tf-idf analysis. Please note that these were the terms that stood out when using the 1-gram analysis out of more than 25,000 tweets and we ignored the term 'unemployment' as that was the name of the hashtag (also proving the basic premise of tf-idf) (Table 1 below).

Table 1. tf-idf analysis—1-gram model

Reason/Location	Effects	Intervention
Michigan	Suffering	Government
Sweden	Pain	Society
Coronavirus	Worrying	Self Employed
Economy	Poverty	Qualify/Eligible
Recession	Struggles	Claims
Layoffs	Depression	
Businesses	Insurance	
China		

So, interestingly, despite the hashtag being in English, we see that the majority of the users referred to Michigan, China and Sweden. Cross-checking against the unemployment data from US during the month of April, Michigan was indeed one of the hardest hit states. The data from China and Sweden

were not available but the frequency of the term indicates the presence of non-English speakers on the hashtag. This goes to the RQ1 above and is consistent with the finding of [10] where the authors argued that social media has played a role in propagation of language—English in this case. Furthermore, we note that coronavirus was the biggest term that appeared as the cause of the unemployment.

Continuing to answer RQ1, looking at the terms related to the effects of unemployment, we see that there are three major concerns, namely financial strain, possible worry about health insurance and the onslaught of depression. While we cannot say for certainty the term ‘insurance’ refers to ‘health insurance’, the context of unemployment indicates that this might be the case.

Looking at RQ2, we can see from above that the social media discusses all the three tenets of traditional research on unemployment, namely causes, effects and possible interventions—in this case, the users look up to the government and society to provide the necessary means. Below, we provide some of the sample tweets that reflect the results obtained above (Table 2).

Table 2. Sample Tweets—1-gram model.

Tweets
Nearly a quarter of #Michigan workers have filed for #unemployment— https://t.co/QzXOHLliQm RT @jenssen_jonas: #Sweden has many more #deaths but much less #unemployment... Look this... https://t.co/QzXOHLliQm The coronavirus pandemic could push half a billion people into poverty, according to @Oxfam Via @wef 6.6 million Americans filed for #unemployment due to the spike in #layoffs and #furloughs. @kayvanwey spoke with @businessinside ... As #unemployment goes up from #COVID-19 so will the consequences of mental illness such as anxiety, depression, drug use & ... Economists expect the U.S. unemployment to soar to a post-Depression record, followed by a recovery that will be moderate a ... Do you know someone who needs the #VRS? #vulnerability #elderly #disabled #depression #addiction #unemployment https://t.co/6MOhiKQmP2 Rising #Unemployment from virus fight threatens #China’s poverty targets https://t.co/klQsW3oiVo Men’s Rights NGO Urges Delhi High Court To Suspend All Maintenance Orders For 3-Months Amidst #Lockdown @narendramodi has likely set in motion #India’s if the not the World’s worst ever disaster, a rampant #covid19 epid ... https://t.co/rfVb6L8cmg #WellsFargo \$WFC discussing their \$4b in loan provisions, and their expectation for the #economic pain ahead I can’t stop thinking, and worrying, about the loss of #jobs due to the #CoronaVirus, and the enormous pain/suffering that co ... BREAKING: Coronavirus: One tenth of US workforce now unemployed as government struggles to process benefits: A US j ... https://t.co/B2mPlvI2Bk

4.2. 2-Gram

To further refine the above, we look at the results obtained from the 2-gram model (Table 3).

Table 3. 2-Gram model.

Reason/Location	Effects	Intervention
coronavirus unemployment	pain suffering	qualify assistance
yemen americans	americans struggle	unemployment coverage/receiving unemployment
rick scott	gig workers	stimulus payments
	lose healthcare	stop thinking
	jobless claims	teleworking paid
	suffering depression	working teleworking
	worrying loss	paid leave
	provide food	compensation law
	thinking worrying	

Once again, we see coronavirus is being used with unemployment indicating that it is being singled out as the reason for unemployment in the US. Furthermore, continuing the trend of Michigan, the bigram ‘yemen americans’ is used extensively—a community mostly found in Michigan. Moreover, the term ‘rick scott’—senator from Florida appeared quite a bit and was filtered in the 1-gram model as the terms ‘rick’ and ‘scott’ individually did not mean much. Looking at the effects, the bigram ‘pain suffering’ occurred very highly but remains ambiguous in terms of physical pain versus the financial pain. Again, looking at the effects, we see that depression was the major issue. The intervention column has more information in the bigram model as users mostly discuss ways to ease the financial suffering. However, one bigram ‘stop thinking’ offered a possible intervention for worrying and depression

as we can see from the Effects column. The Table 4 provides sample tweets reflecting the results of 2-gram analysis.

Table 4. Sample Tweets—2-gram model.

Tweets
I can't stop thinking, and worrying, about the loss of #jobs due to the #CoronaVirus, and the enormous pain/suffering that co ... #Yemen
Americans struggle to provide food for their families because of an increase in #unemployment which is being blamed o ... #GOP #Voters
Here is what .@SenRickScott Senator Rick Scott had to say about the #Florida #Unemployment system. ... https://t.co/oceHS3bm20 The NJDOL has issued unemployment insurance guidance for the self-employed, independent contractors and gig workers https://t.co/... ... Do I Qualify for Assistance? If you are not currently working, teleworking or on paid leave, you are eligible for one of ...
Questions about Unemployment Compensation due to COVID-19? Check out @CommLegalAid 's Ask A Lawyer series tomorrow ... @StatistaCharts: Almost 17mn Americans have filed #jobless claims in the last 3 weeks, over 2mn in #California alone. These are levels o ... Do I Qualify for Assistance? If you are not currently working, teleworking or on paid leave, you are eligible for ...
@narendramodi has likely set in motion #India's if the not the #World's worst ever disaster, a rampant #covid19 epid ... You have questions about #unemployment benefits, #stimulus payments and more. Tonight at 6, @CBSDFW gets you answers ... I can't stop thinking, and worrying, about the loss of #jobs due to the #CoronaVirus, and the enormous pain/suffering that co ... #COVID-19 Information: #StimulusPayments to individuals, expanded #unemployment coverage contained in the

4.3. 3-Gram

Finally, we look at the 3-gram results from our analysis. In terms of location, we see that Texas workforce has been added to the mix. The effects were similar to what we found earlier as is also the case for the intervention mechanisms. While no new results stand out in 3-gram model for effects and intervention, we see that the results below confirm the results we got in both the 1-gram and 2-gram model (Table 5).

Table 5. 3-Gram model.

Reason/Location	Effects	Intervention
lost job COVID-19 yemen americans struggle annette_taddeo rick scott txworkforce qualify assistance	COVID-19 apply healthcare million people filed americans filed unemployment americans struggle provide families increase unemployment enormous pain suffering coronavirus enormous pain worrying loss jobs provide food families	Expanded unemployment coverage sweeping unemployment compensation get stimulus payments/stimulus payments individuals teleworking paid leave currently working teleworking paid leave eligible stop thinking worrying

Looking at the 3-gram model, we see a confirmation of results from the 2-gram model. As opposed to the case of 1-gram vs 2-gram where certain words could have ambiguous meaning, e.g., great depression vs. suffering depression. The following sample tweets confirm that not much changed in terms of results from the 2-gram model to the 3-gram model (Table 6).

Table 6. Sample Tweets—3-gram mode.

Reason/Location
I can't stop thinking, and worrying, about the loss of #jobs due to the #CoronaVirus, and the enormous pain/suffering that co ... #Yemen
@Annette_Taddeo: "Rick Scott a decade ago instigated a sweeping #unemployment compensation law designed to punish workers who lost their ... More than 6.6 million people filed for #unemployment in the past week. 10 million in two weeks. Now, @StevenMnuchin1 an ... @TXWorkforce: Do I Qualify for Assistance? If you are not currently working, teleworking or on paid leave, you are eligible for one of ...
Image of hardship: Arturo looking for help but #unemployment office closed. The #SJ machinist lost job when shop closed due ... If you have lost your job during #COVID-19 you can apply for healthcare on https://t.co/tZEsq3K9uP up to 60 days before or ... Americans struggle to provide food for their families because of an increase in #unemployment which is being blamed o ... "Rick Scott a decade ago instigated a sweeping #unemployment compensation law designed to punish workers who lost their ... Okay, so aren't the stimulus payments for individuals supposed to be retroactive? From my understanding to either 3/28 or 4/4? Eith ... @RepRashida: TOMORROW: I'll be holding a tele- #townhall on understanding and navigating expanded #unemployment benefits and leave polic ...

4.4. Hashtags

Lastly, we also looked at the hashtags that were mentioned in the tweets. While there were more than 2500 hashtags that we encountered in our analysis, the following ten hashtags (Table 7) were mentioned more than 100 times.

Table 7. Top Hashtags.

Top Hashtags
#economy
#unemployment
#coronavirus
#COVID-19
#florida
#stimulus
#sweden
#jobs
#yemenamericans
#texas

Looking at the hashtags above, we see that the results confirm our earlier results and, more importantly, we see that the results we obtain confirm the findings of traditional research on unemployment where users confirm the importance of intervention when it comes to involuntary unemployment.

5. Conclusions

The main question that motivated our work was: can we depend on data from social media as a basis for deducing meaningful results? To do so, we looked at the results mentioned in the literature when it comes to unemployment. The choice of topic was driven by the fact that it has been studied from various angles over many decades and researchers have presented plethora of studies covering social and health aspects of unemployment. Furthermore, many studies in the last decades have presented conclusions solely based on social media data. The implicit underlying basis of such work is rooted in the concept of wisdom of crowd, where the argument is that the decision taken by a group leads to an optimum solution [63–65]. However, it behooves us to also look at the fallacy of the crowd concept, which contradicts the underlying assumption of wisdom of the crowd as described in [66–68]. Traditionally, gathering statistics would require clearly defining various control variables such as ethnicity, age, nationality, language, etc. Social media on the other hand does not explicitly provide such delineations and hence we ask: can we depend on the volume of the data from social media/microblogging to ensure the lack of bias in the conclusions drawn from such data?

Keeping the above motivation in mind, our results replicated the traditional model when it comes to unemployment. Specifically, we found three tenets of prior research on unemployment reflected in data gathered from Twitter despite the difference of time and cultural differences. Given that the tweets gathered were in English, we found that the majority of results reflected the community in USA. We plan to extend this work by looking at other hashtags gathered from this work and compare to our results. It would also be interesting to compare the results of this study to another language. Lastly, it remains to be seen whether crowdsourcing techniques can be refined enough to use various segmentation techniques, such as age, ethnicity, etc., that play a big role in medical research. Our current work looked at the Twitter data as a whole rather than segmenting it in a traditional fashion. This is a shortcoming we intend to address in the future.

Author Contributions: S.A.: Conceptualization, Sections 1, 2 and 5 and writing—original draft preparation and software; A.E.R.: methodology, Sections 3 and 4 and writing—original draft preparation and software; A.S.: Section 4, validation and software and writing—review and editing; A.A. (Asma Aljaberi), O.A. and A.A. (Abrar Alsheraigi): investigation, data curation, Sections 2.2 and 3. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Coppersmith, G.; Harman, C.; Dredze, M. Measuring post traumatic stress disorder in Twitter. In Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, Ann Arbor, MI, USA, 1–4 June 2014; The AAAI Press: Menlo Park, CA, USA, 2014; pp. 579–582.
2. Linn, M.W.; Sandifer, R.; Stein, S. Effects of unemployment on mental and physical health. *Am. J. Public Health* **1985**, *75*, 502–506. [[CrossRef](#)] [[PubMed](#)]
3. Corrigan, P.W.; Miller, F.E. Shame, blame, and contamination: A review of the impact of mental illness stigma on family members. *J. Mental Health* **2004**, *13*, 537–548. [[CrossRef](#)]
4. Rüsçh, N.; Müller, M.; Ajdacic-Gross, V.; Rodgers, S.; Corrigan, P.; Rössler, W. Shame, perceived knowledge and satisfaction associated with mental health as predictors of attitude patterns towards help-seeking. *Epidemiol. Psychiatr. Sci.* **2014**, *23*, 177–187. [[CrossRef](#)] [[PubMed](#)]
5. Eales, M.J. Shame among unemployed men. *Soc. Sci. Med.* **1989**, *28*, 783–789. [[CrossRef](#)]
6. Rantakeisu, U.; Starrin, B.; Hagquist, C. Unemployment, shame and ill health—An exploratory study. *Scand. J. Soc. Welf.* **1997**, *6*, 13–23. [[CrossRef](#)]
7. Rantakeisu, U.; Starrin, B.; Hagquist, C. Financial hardship and shame: A tentative model to understand the social and health effects of unemployment. *Br. J. Soc. Work* **1999**, *29*, 877–901. [[CrossRef](#)]
8. Lippmann, S. Rethinking risk in the new economy: Age and cohort effects on unemployment and re-employment. *Hum. Relat.* **2008**, *61*, 1259–1292. [[CrossRef](#)]
9. Parameswaran, M.; Whinston, A.B. Social computing: An overview. *Assoc. Inf. Syst.* **2007**, *16*, 762–780. [[CrossRef](#)]
10. Thorne, S.L. Artifacts and Cultures-of-Use in Intercultural Communication. *Lang. Learn. Technol.* **2003**, *7*, 38–67. Available online: https://scholarspace.manoa.hawaii.edu/bitstream/10125/25200/07_02_thorne.pdf (accessed on 27 September 2020).
11. Dunkels, E. Bridging the Distance: Children’s Strategies on the Internet. Ph.D. Thesis, Group of Interactive Media and Learning, Umeå University, Umeå, Sweden, 2007.
12. Savin, D.; Glueck, D.A.; Chardavoigne, J.; Yager, J.; Novins, D.K. Bridging cultures: Child psychiatry via videoconferencing. *Child. Adolesc. Psychiatr. Clin. N. Am.* **2011**, *20*, 1125–1134. [[CrossRef](#)]
13. Youyou, W.; Kosinski, M.; Stillwell, D. Computer-based personality judgments are more accurate than those made by humans. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 41036–41040. [[CrossRef](#)] [[PubMed](#)]
14. Paniagua, J.; Korzynski, P. *Social Media Crowdsourcing*; Springer: Berlin/Heidelberg, Germany, 2017.
15. Jordan, S.E.; Hovet, S.E.; Fung, I.C.; Liang, H.; Fu, K.W.; Tse, Z.T.H. Using twitter for public health surveillance from monitoring and prediction to public response. *Data* **2019**, *4*. [[CrossRef](#)]
16. Schvaneveldt, R.W.; Meyer, D.E. Lexical ambiguity, semantic context, and visual word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* **1976**, *2*, 243–256. [[CrossRef](#)] [[PubMed](#)]
17. Demszky, D.; Garg, N.; Voigt, R.; Zou, J.; Gentzkow, M.; Shapiro, J.; Jurafsky, D. Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings. *arXiv* **2019**, arXiv:1904.01596.
18. Rajput, A.; Ahmed, S. Making A Case for Social Media Corpus for Detecting Depression. *arXiv* **2019**, arXiv:1902.00702.
19. Berardis, D.D.; Fornaro, M.; Orsolini, L.; Valchera, A.; Carano, A.; Vellante, F.; Perna, G.; Serafini, G.; Gonda, X.; Pompili, M.; et al. Alexithymia and suicide risk in psychiatric disorders: A mini-review. *Front. Psychiatry* **2017**, *8*, 148. [[CrossRef](#)]
20. Mroz, T.; Savage, T.H. The Long-Term Effects of Youth Unemployment. *J. Human Resour.* **2006**, *41*, 259–293. Available online: <https://EconPapers.repec.org/RePEc:uwp:jhriss:v:41:y:2006:i:2:p259-293> (accessed on 27 September 2020). [[CrossRef](#)]
21. Arulampalam, W. Is Unemployment Really Scarring? Effects of Unemployment Experiences on Wages. *Econ. J.* **2001**, *111*, F585–F606. [[CrossRef](#)]
22. Kessler, R.C.; Turner, J.B.; House, J.S. Effects of unemployment on health in a community survey: Main, modifying, and mediating effects. *J. Soc. Issues* **1988**, *44*, 69–85. [[CrossRef](#)]
23. Mastekaasa, A. Unemployment and Health: Selection Effects. *J. Community Appl. Soc. Psychol.* **1996**, *6*, 189–205. [[CrossRef](#)]

24. Dooley, D.; Fielding, J.; Levi, L. Health and Unemployment. *Annu. Rev. Public Health* **1996**, *17*, 1449–1465. [[CrossRef](#)] [[PubMed](#)]
25. Pohlen, L. Unemployment and social exclusion. *J. Econ. Behav. Organ.* **2019**, *164*, 273–299. [[CrossRef](#)]
26. Voßemer, J.; Gebel, M.; Täht, K.; Unt, M.; Högberg, B.; Strandh, M. The Effects of Unemployment and Insecure Jobs on Well-Being and Health: The Moderating Role of Labor Market Policies. *Soc. Indic. Res.* **2018**, *138*, 31229–31257. [[CrossRef](#)]
27. Mihaela, C.E. *The Education—An Important Factor on Unemployment And Profession*; Constantin Brâncuși, University of Târgu Jiu: Târgu Jiu, Romania, 2013; pp. 219–226, Issue 6.
28. Dixit, J.; Pankaj, T.; Gupta, S.; Singh, P.; Gupta, H. Educated Unemployed: A Challenge before Sustainable Education. In Proceedings of the International Conference on Sustainable Manufacturing: Issues, Trends and Practices (ICSM-2011), BITS, Pilani, India, 10–12 November 2011.
29. Marcus, J. The effect of unemployment on the mental health of spouses—Evidence from plant closures in Germany. *J. Health Econ.* **2013**, *32*, 3546–3558. [[CrossRef](#)] [[PubMed](#)]
30. Engel-Yeger, B.; Caterina, M.; Giorgio, R.; Paola, S.; Pierre, A.; Maurizio, P.; Mario, A.; Gianluca, S. Extreme sensory processing patterns and their relation with clinical conditions among individuals with major affective disorders. *Psychiatry Res.* **2016**, *236*, 112–118. [[CrossRef](#)]
31. Kunze, L.; Suppa, N. Bowling alone or bowling at all? The effect of unemployment on social participation. *J. Econ. Behav. Organ.* **2017**, *133*, C213–C235. [[CrossRef](#)]
32. Suphan, A.; Feuls, M.; Fieseler, C.; Meckel, M. The Supportive Role of Social Media Networks for those Out of Work. In Proceedings of the 2013 46th Hawaii International Conference on System Sciences, Wailea, Maui, HI, USA, 7–10 June 2013; pp. 3312–3321.
33. Shahkooh, K.A.; Azadnia, M.; Shahkooh, S.A. An Investigation into the Effect of Information Technology on the Rate of Unemployment. In Proceedings of the 2008 Third International Conference on Convergence and Hybrid Information Technology, Busan, Korea, 11–13 November 2008; Volume 1, pp. 61–65. [[CrossRef](#)]
34. Strawn, G. IT and Future Unemployment: Part 2. *IT Prof.* **2017**, *19*, 170–172. [[CrossRef](#)]
35. Proserpio, D.; Counts, S.; Jain, A. The Psychology of Job Loss: Using social Media Data to Characterize and Predict Unemployment. In Proceedings of the 8th ACM Conference on Web Science, Hannover, Germany, 22–25 May 2016; ACM: New York, NY, USA, 2016; pp. 223–232. [[CrossRef](#)]
36. Mincer, J. Unemployment Effects of Minimum Wages. *J. Polit. Econ.* **1976**, *84*, S87–S104. [[CrossRef](#)]
37. Ryu, P.-M. Predicting the Unemployment Rate Using Social Media Analysis. *J. Inf. Process. Syst.* **2018**, *14*, 4904–4915. [[CrossRef](#)]
38. Bokányi, E.; Lábszki, Z.; Vattay, G. Prediction of employment and unemployment rates from Twitter daily rhythms in the US. *EPJ Data Sci.* **2017**, *6*, 14. [[CrossRef](#)]
39. Nirmala, C.; Roopa, G.; Kumar, K.N. Twitter data analysis for unemployment crisis. In Proceedings of the 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Davangere, Karnataka, India, 29–31 October 2015; pp. 420–423.
40. Antenucci, D.; Cafarella, M.; Levenstein, M.; Ré, C.; Shapiro, M.D. *Using Social Media to Measure Labor Market Flows*; Work Paper 20010; National Bureau of Economic Research: Cambridge, MA, USA, 2014. [[CrossRef](#)]
41. Subrahmanian, V.D.; Adali, S.; Brink, A.; Emery, R.; Lu, J.J.; Rajput, A.; Rogers, T.J.; Ross, R.; Ward, C. HERMES: A Heterogeneous Reasoning and Mediator System. 1995. Available online: <http://www.cs.umd.edu/projects/hermes/publications/postscripts/tois.ps> (<http://www.cs.umd.edu/projects/hermes/publications/postscripts/tois.ps> (accessed on 29 September 2020)).
42. Rajput, A.; Rotenstreich, S. Making A Case for Resource Management in a P2P Environment. In Proceedings of the International Conference on Information and Knowledge Engineering. IKE'04, Las Vegas, NV, USA, 21–24 June 2004.
43. Wazny, K. Crowdsourcing' ten years in: A review. *J. Glob. Health* **2017**, *7*, 020602. [[CrossRef](#)] [[PubMed](#)]
44. Estellés-Arolas, E.; González-Ladrón-De-Guevara, F. Towards an integrated crowdsourcing definition. *J. Inf. Sci.* **2012**, *38*, 189–200. [[CrossRef](#)]
45. Ranard, B.L.; Ha, Y.P.; Meisel, Z.F.; Asch, D.A.; Hill, S.S.; Becker, L.B.; Seymour, A.K.; Merchant, R.M. Crowdsourcing—harnessing the masses to advance health and medicine, a systematic review. *J. Gen. Intern. Med.* **2014**, *29*, 187–203. [[CrossRef](#)] [[PubMed](#)]
46. Tucker, J.D.; Day, S.; Tang, W.; Bayus, B. Crowdsourcing in medical research: Concepts and applications. *PeerJ* **2019**, *7*, e6762. [[CrossRef](#)] [[PubMed](#)]

47. Doan, A.; Ramakrishnan, R. Crowdsourcing systems on the World-Wide Web. *Commun. ACM* **2011**, *54*. [[CrossRef](#)]
48. Rajput, A.E.; Sarirete, A.; Desouky, T.F. Using Crowdsourcing to Identify a Proxy of Socio-economic Status. In *The International Research & Innovation Forum*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 479–486.
49. Ahmed, S.; Rajput, A.; Sarirete, A.; Chaudhery, T. Social Media Platform: Measuring Readability and Socio-Economic Status. *Preprint* **2020**. [[CrossRef](#)]
50. Llorente, A.; Garcia-Herranz, M.; Cebrian, M.; Moro, E. Social Media Fingerprints of Unemployment. *PLoS ONE* **2015**, *10*, e0128692. [[CrossRef](#)]
51. Ahmed, S.; Rajput, A.; Sarirete, E.A.; Bahwireth, R.; Almeahadi, A.; Khimi, W. Characterizing Female Workplace Bullying via Social Media. *Preprint* **2020**. [[CrossRef](#)]
52. Rajput, A.; Brahimi, T. Characterizing internet of medical things/personal area networks landscape. In *Innovation in Health Informatics*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 353–371.
53. Murdoch, T.B.; Detsky, A.S. The inevitable application of big data to health care. *JAMA* **2013**, *309*, 1351–1352. [[CrossRef](#)]
54. Chen, E.E.; Wojcik, S.P. A practical guide to big data research in psychology. *Psychol. Methods* **2016**, *21*, 458–474. [[CrossRef](#)]
55. Monteith, S.; Glenn, T.; Geddes, J.; Bauer, M. Big data are coming to psychiatry: A general introduction. *Int. J. Bipolar Disord.* **2015**, *3*, 21. [[CrossRef](#)] [[PubMed](#)]
56. Rajput, A.E.; Ahmed, S.M. Big data and social/medical sciences: State of the art and future trends. *arXiv* **2019**, arXiv:1902.00705.
57. Choudhury, M.D. Social Media for Mental Illness Risk Assessment, Prevention and Support. In Proceedings of the 1st ACM Workshop on Social Media World Sensors, Guzelyurt, Northern Cyprus, 1 September 2015. [[CrossRef](#)]
58. Choudhury, M.D.; Gamon, M.; Counts, S.; Horvitz, E. Predicting Depression via Social Media. July 2013. Available online: <https://www.microsoft.com/en-us/research/publication/predicting-depression-via-social-media/> (accessed on 29 September 2020).
59. Ahmed, S.M.; Rajput, A. Threats to Patients' Privacy in Smart Healthcare Environment. *Innov. Health Inform.* **2020**. [[CrossRef](#)]
60. Ahmed, S. BYOD, personal area networks (PANs) and IOT: Threats to Patients Privacy. In *The International Research & Innovation Forum*; Visvizi, A., Lytras, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2019. [[CrossRef](#)]
61. Jones, K.S. A statistical interpretation of term specificity and its application in retrieval. *J. Doc.* **1972**, *28*, 11–21. [[CrossRef](#)]
62. Rajput, A. Natural Language Processing, Sentiment Analysis, and Clinical Analytics. *Innov. Health Inform.* **2020**. [[CrossRef](#)]
63. Bechter, C.; Jentzsch, S.; Frey, M. From wisdom to wisdom of the crowd and crowdfunding. *J. Commun. Comput.* **2011**. forthcoming. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1851142 (accessed on 29 September 2020).
64. Lüttgens, D.; Pollok, P.; Antons, D.; Piller, F. Wisdom of the crowd and capabilities of a few: Internal success factors of crowdsourcing for innovation. *J. Bus. Econ.* **2014**, *84*, 339–374. [[CrossRef](#)]
65. Kremer, I.; Mansour, Y.; Perry, M. Implementing the 'wisdom of the crowd'. *J. Polit. Econ.* **2014**, *122*, 988–1012. [[CrossRef](#)]
66. Allport, F.H. The group fallacy in relation to social science. *Am. J. Sociol.* **1924**, *29*, 688–706. [[CrossRef](#)]
67. Follett, M.P. *The Crowd Fallacy*; Longmans, Green and Co.: New York, NY, USA, 1923.
68. Castellion, G.; Markham, S.K. Perspective: New Product Failure Rates: Influence of A rgumentum ad Populum and Self-Interest. *J. Prod. Innov. Manag.* **2013**, *30*, 976–979. [[CrossRef](#)]

