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

Samara Ahmed¹, Adil E. Rajput²*, Akila Sarirete³*, Asma Aljaberib, Oohoud Alghanem⁵ and Abrar Alsheraigib⁶

¹ Psychiatry Division, College of Medicine, King Abdulaziz University, Jeddah, KSA; samraa2018@gmail.com
² Computer Science Department, College of Engineering, Effat University, Jeddah, KSA; aailahi@effatuniversity.edu.sa
³ Computer Science Department, College of Engineering, Effat University, Jeddah, KSA; asarirete@effatuniversity.edu.sa
⁴ Information Systems Department, College of Engineering, Effat University, Jeddah, KSA; aaljaberib@effat.edu.sa
⁵ Information Systems Department, College of Engineering, Effat University, Jeddah, KSA; Ohalghanem@effat.edu.sa
⁶ Information Systems Department, College of Engineering, Effat University, Jeddah, KSA; aaalsheraigi@effat.edu.sa

* Correspondence: aailahi@effatuniversity.edu.sa

Abstract: Social media, traditionally reserved for social exchanges on the net, 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. In this work, we compare the feedback gathered from social media using crowdsourcing techniques to results obtained prior to the advent of Big Data. We find that the results are consistent in terms of 1) financial strain is the biggest stressor and concern, 2) onslaught of depression is typical and 3) possible interventions including reemployment and support from friends and family is crucial 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 from 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.

Social computing has emerged as a discipline where users employ various computing devices for social use rather than increasing productivity. The work done in [3] points out the change in social interaction among people in the past few years. The social media platforms offer plethora of data in
forms of free text that embodies opinions of users across the globe. Thorne, Steven in [4] provides a comprehensive study where he argues how 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 internet is similar when engaging in non-internet communication. Dunkels in his work [5] did a study on children’s perceived dangers (Swedish 6th graders between the age of 11 and 13) of the net 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. [6] 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 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 [7] 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 allowing outsourcing tasks to humans regardless of their physical location) has been employed increasingly to group relevant data together. Paniagua & Korzynski [8] show how the concept of crowdsourcing is part and parcel of the social media platform. Thus, Twitter platform allowed in an emergency situation to gather feedback from volunteers in the affected region and helped the authorities to respond appropriately [9]. 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 [10].

Demzsky et al. in [11] 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 [12] make a case for a corpus to assist mental health professionals in detecting depression among users provided some group of people. The researchers make use of 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 platform provide sufficient data that conforms 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 data 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 [13], 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 [14] 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 [15] 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 [16] took an alternative view in literature where he argues, based on a study in Norway, that mentally health 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 [17] 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 [18] confirm 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. [19] takes a step further and limit the effects of unemployment to mental health and...
well-being but not on physical health. The work done in [20] 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 [21] in India. Using data from the German Socio-Economic Panel Study (SOEP) from 2002 through 2010 found the negative effects of unemployment spouse’s mental health and the fact that unemployment had severe consequences on both the unemployed and their spouse [22].

2.2. Social media and Unemployment

Kunze & Suppa in [23] concluded that unemployment affects negatively on 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 [24]. [21] 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 [25] [26, p. 2] explore outsourcing of jobs in general and IT in specific to see the effects on unemployment. Proserpio et al. in [27] did a very thorough study of 230,000 U.S 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. [24] 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 [28] 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 [29] used social media 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 [27] 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 [30] 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 [31], [11], [32] 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 heterogeneous data from various sources [33]. The advent of Peer-to-Peer networks also were a step forward to conflate data in unstructured form [34]. Wazny in [35] 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 medical field as it has the potential to gather vast amount of relevant data. Doan et. al., in [36] discussed the
use of 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 SES [37],[38],[39] use the same concept to predict the socioeconomic status. The work done in [12] uses a similar technique to detect depression but the work is more focused on building corpus.

2.4. Big Data/Social Media and Mental Health

Murdoch and Detsky [40] introduced the need for applying the big data techniques to medical field to gather better insights. Chen & Wojcik described a framework on applying big data in the field of psychology and mental health [41]. 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 [42] while the work done in [43] provides an overview of work done in medical sciences along with various techniques that are employed. Dechoudhry et. al. has laid foundation for work in applying such techniques to social media platform [44],[45]. 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 to ensure that we address the privacy concerns of using such data [46],[47]. 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:

1. Collected over 25000 tweets under the hashtag #unemployment
2. We use the nltk toolkit to parse the texts and get rid of the stop-words (recurring words such as articles that need to be filtered out)
3. We use the t-f-idf algorithm [48] to generate the keywords
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 and the n-gram model is not gappy.
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.
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 a 16G RAM

3.1.1. The 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 [49] for pertinent details.

4. Sklearn: This library helps in big data analysis such as classification, regression, clustering etc.

4. Results and Discussion

As a reminder, we restate the three questions we posed as the goals of this study

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 data 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?

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 COVID19 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>
<thead>
<tr>
<th>Reason/Location</th>
<th>Effects</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michigan</td>
<td>Suffering</td>
<td>Government</td>
</tr>
<tr>
<td>Sweden</td>
<td>Pain</td>
<td>Society</td>
</tr>
<tr>
<td>Coronavirus</td>
<td>Worrying</td>
<td>Self Employed</td>
</tr>
<tr>
<td>Economy</td>
<td>Poverty</td>
<td>Qualify/Eligible</td>
</tr>
<tr>
<td>Recession</td>
<td>Struggles</td>
<td>Claims</td>
</tr>
<tr>
<td>Layoffs</td>
<td>Depression</td>
<td></td>
</tr>
<tr>
<td>Businesses</td>
<td>Insurance</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

So interestingly despite the hashtag was in English, we see that 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 was 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 [4] 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 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.

4.2. 2-gram

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

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

Once again, we see coronavirus is being used with unemployment indicating that it is being singled out as the reason for unemployment in US. Furthermore, continuing the trend of Michigan, the bigram ‘yemen americans’ is used extensively – a community mostly found in Michigan. Also, 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.

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 1-gram and 2-gram model (Table 3).
<table>
<thead>
<tr>
<th>Reason/Location</th>
<th>Effects</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>lost job covid19</td>
<td>covid19 apply healthcare</td>
<td>Expanded unemployment coverage</td>
</tr>
<tr>
<td>yemen americans struggle</td>
<td>million people filed</td>
<td>sweeping unemployment compensation</td>
</tr>
<tr>
<td>annette_taddeo rick scott</td>
<td>americans filed</td>
<td>get stimulus payments/stimulus payments individuals</td>
</tr>
<tr>
<td>txworkforce qualify assistance</td>
<td>americans struggle provide</td>
<td>teleworking paid leave</td>
</tr>
<tr>
<td></td>
<td>families increase unemployment</td>
<td>currently working teleworking</td>
</tr>
<tr>
<td></td>
<td>enormous pain suffering</td>
<td>paid leave eligible</td>
</tr>
<tr>
<td></td>
<td>coronavirus enormous pain</td>
<td>stop thinking worrying</td>
</tr>
<tr>
<td></td>
<td>worrying loss jobs provide</td>
<td>provide food families</td>
</tr>
</tbody>
</table>

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 were mentioned more than 100 times.

Table 4. Top Hashtags

<table>
<thead>
<tr>
<th>Top Hashtags</th>
</tr>
</thead>
<tbody>
<tr>
<td>#economy</td>
</tr>
<tr>
<td>#unemployment</td>
</tr>
<tr>
<td>#coronavirus</td>
</tr>
<tr>
<td>#covid19</td>
</tr>
<tr>
<td>#florida</td>
</tr>
<tr>
<td>#stimulus</td>
</tr>
<tr>
<td>#sweden</td>
</tr>
<tr>
<td>#jobs</td>
</tr>
<tr>
<td>#yemenamericans</td>
</tr>
<tr>
<td>#texas</td>
</tr>
</tbody>
</table>

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

In this paper, we put together a framework to scavenge data on involuntary unemployment and people’s reaction to it. Comparing to the traditional research on unemployment, we established that 1) the financial strain is the most difficult part of the involuntary unemployment, 2) it causes mental health issue specially depression and 3) people look for intervention both in terms of reemployment and family/friends’ support. We built upon the earlier work done on active and passive crowdsourcing and gathered over 25,000 tweets under the hashtag ‘#unemployment’. Using the tf-idf algorithm, we looked at n-gram models for n = 1, 2 and 3 while also gathering over 2500 hashtags and distilled it into ten hashtags that were mentioned more than 100 times. Looking at the data from social media, our results replicated the traditional model and we found the three tenets of prior research on unemployment. We plan to extend this work by looking at other hashtags gathered from this work and compare to our results. What will also be interesting would be to compare the results gathered in English language to another language.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used “Samara Ahmed: Conceptualization, 1,2 and writing—original draft preparation and software; Adil Rajput: methodology, 3, 4 and writing—original draft preparation and software; Akila Sarirete: 4, validation and software and writing—review and editing; Asma Aljaberi, Ohoud Alghanem and Abrar Alsheraigi: investigation, data curation, 2, 2 and 3.

Funding: “This research received no external funding”.

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

References


