

COVID-19: Understanding Construction Industry Responses on Twitter in the Emergence of Novel Coronavirus

Priyanka Linge¹; M. Ahmed Rusho²; Md. Ashraf Ahmed³; and Arif Mohaimin Sadri, Ph.D.⁴

¹M.Sc. Student, School of Computing and Information Sciences, FIU College of Engineering and Computing, Miami, FL. Email: pling004@fiu.edu

²M.Sc. Student, Moss Dept. of Construction Management, FIU College of Engineering and Computing, Miami, FL. Email: mrush009@fiu.edu

³Ph.D. Candidate, Dept. of Civil and Environmental Engineering, FIU College of Engineering and Computing, Miami, FL. Email: mpave002@fiu.edu

⁴Incoming Assistant Professor, School of Civil Engineering and Environmental Science, Univ. of Oklahoma, Norman, OK. Email: sadri.buet@gmail.com

ABSTRACT

COVID-19 has spread rampantly in 2020 and people in different countries experienced unprecedented challenges to respond immediately. It has deeply shaken different industries and the overall economy. Many businesses were closed during this disease outbreak, but the construction industry continued to operate while others struggled. As the pandemic spread, the intensity of social media interactions dramatically increased as people started maintaining physical distances and staying at home. Recent studies have shown that social media serves the purpose of disseminating information in major crises and such social interactions correlate with what happens in the real world. Unlike traditional media, social media data also offered sufficient statistical power to understand risk communication patterns. The construction industry adopted innovative strategies to continue its operations and activities while responding to a number of concerns associated with health risks of the workers on site. This study aims at identifying the emerging communication patterns of construction stakeholders on social media (Twitter) during COVID-19. Around 12 million tweets were collected in the early days of the ongoing pandemic (March, April, and May 2020) by Twitter Application Programming Interface (API), in the region of North America; then filtered this data based on keywords pertaining to COVID-19 and construction activities. Such information was processed by applying several machine learning and natural language processing techniques. The spatiotemporal patterns of construction risk communication revealed in this study would support policymakers, businesses, and other industries to take more efficient measures and be more resilient both in the current and future pandemics.

INTRODUCTION AND MOTIVATION

Coronavirus disease (COVID-19) caused by the spread of SARS-coV-2 infection which became an epidemic in China from December 2019. It was announced by World Health Organization (WHO) on 11 March 2020 as a global pandemic (Mojumder et al. 2021). The explanation behind this uncontrollable spread of the disease is that by means of cough sneeze or even talk droplets, the infected person may not display the symptoms for days and may infect others. The virus stays active for hours to approximately 3 days on the surface of variable materials. It has spread to more than 200 countries and this pandemic is affecting more than 175 million people, of which about 3.8 million have died. Out of the 175 million more than 34 million cases

have been found in USA causing around 615,000 deaths until June 2021 (Worldometer 2020). Given this scenario it's important to examine each aspect of this pandemic closely.

Not only has this pandemic impacted the lives of individuals, but it has also had disruptive effects on the economy. The novel coronavirus (COVID-19) pandemic outbreak has already put a huge strain on the resources and economies of many countries around the world. Professionals in healthcare were overwhelmed. People were forced to stay at home, and several major meetings were cancelled, including sporting competitions, religious and cultural gatherings. Many schools have announced that they will be closed for this full year and that they will proceed online. Many industries struggled to function during this pandemic even the construction industry slowed down but slowly it kept on operating (Knight 2020). The construction industry being one of the essential businesses was functioning and adapted to the pandemic (CBO 2020).

This study will help policymakers to be aware of the social media communication trends by the construction industry so that in such pandemic scenarios they can get direction for successful communication in the future. Over a span of three months, we collected tweets in the region of North America. In order to analyze user feelings and concerns over time-based on text, geo-locations, and timestamps of such tweets, we have applied several machine learning techniques (topic model) and data mining approaches (i.e., most frequent words, Bi-gram).

BACKGROUND AND RELATED WORK

The outbreak of COVID-19 is recent and outrageous. Many researchers worked on data related to Covid-19. Haleem et al. explained in depth how various areas such as healthcare, economy, and society have been affected by this pandemic (Haleem et al. 2020). Ceukelaire et al. provide a report on how various countries have reacted to the pandemic thus far and what the world might learn from this pandemic to help develop a better public health system. While doing so, authors have also illustrated what factors in other parts of the world influence health systems and shown the path to stronger health systems (De Ceukelaire and Bodini 2020).

Broniatowski et al. attempted to equate the credibility of COVID-19 tweets with other datasets related to health (Broniatowski et al. 2020). In order to investigate the performance of crisis communication, Ahmed et al. analyzes the tweets of public health organizations and other influencing actors (Ahmed et al. 2020). Camargo discusses the rapid rise in the number of COVID-19 discourses and the challenges to science and public policy due to the overload of knowledge (Camargo Jr 2020). Singh et al. discusses the amount of communication taking place on social media about covid-19 and the quality of information based on the shared URL links (Singh et al. 2020). An updated epidemic model called the controlled-SIR model was introduced by Gros et al. Using the model, the rate of reproduction of diseases is shown to evolve dynamically in response to social and political reactions (Gros et al. 2020).

Bagheri et al. used Twitter data to conduct sentiment analysis on two topics, McDonald's and KFC related tweets were gathered to demonstrate which restaurant is more popular (Bagheri and Islam 2017). Sapul et al. compared the results of k-means, CLOPE clustering, and Latent Dirichlet Allocation (LDA) algorithms to detect trending topics on the Twitter platform (Sapul et al. 2017). Sadri et al. studied a large number of tweets to determine the communication pattern for individuals affected by one of the deadliest hurricanes named Sandy. The author highlighted the most popular terms and various topics created based on these tweets in this study (Sadri et al. 2018). Another study on hurricane Michael and construction related tweets analyzed user sentiments and concerns over time and space based on text and timestamps (Ahmed et al. 2020).

DATA COLLECTION

Social media has been found to play an important role in multiple crisis communications in recent years. Twitter, a microblog service is considered as a reliable source of information (Castillo et al. 2011) which ensures effective information dissemination during crisis. In this study, we have focused on the tweets related to the construction industry during the COVID-19 pandemic. In order to conduct Twitter data analysis, Twitter Application Programming Interface (API) is used to collect large-scale geo-tagged tweets from North America. The data collection started when the covid was declared as a pandemic by WHO, and it was continued for three months (March 13, 2020 to May 31, 2020). Originally around 12 million tweets were collected using Twitter API and then filtered based on corona-related keywords such as 'covid', 'corona', 'ncov'; and construction-related keywords such as 'construction', 'building', 'construct', 'subcontractor', 'crane', 'concrete', 'soil', 'schedule', 'estimate', and 'cement'. The filtering process (by the mentioned keywords) separated the tweets related to COVID-19 and construction industry from the originally collected Twitter dataset. We obtained over 34,800 tweets about Corona and construction-related topics after filtering the tweets based on the list of keywords.

METHODOLOGY

The Twitter data used in this study was collected from the North American region using the tweepy library in python. All of these tweets were compiled into a single JSON file and filtered for English tweets. NLTK library was used to delete bad words and non-English tweets in the pre-processing of bad words. The NLTK-Natural Language Toolkit helps to work with human language data and offers quick interfaces for organizational and lexical tools along with tokenization, classification, etc. text processing libraries. Later, the CSV file was created by extracting important variables such as user ID- containing the unique ID assigned to each user using the Twitter platform, Tweet ID- unique ID assigned to each user tweet.

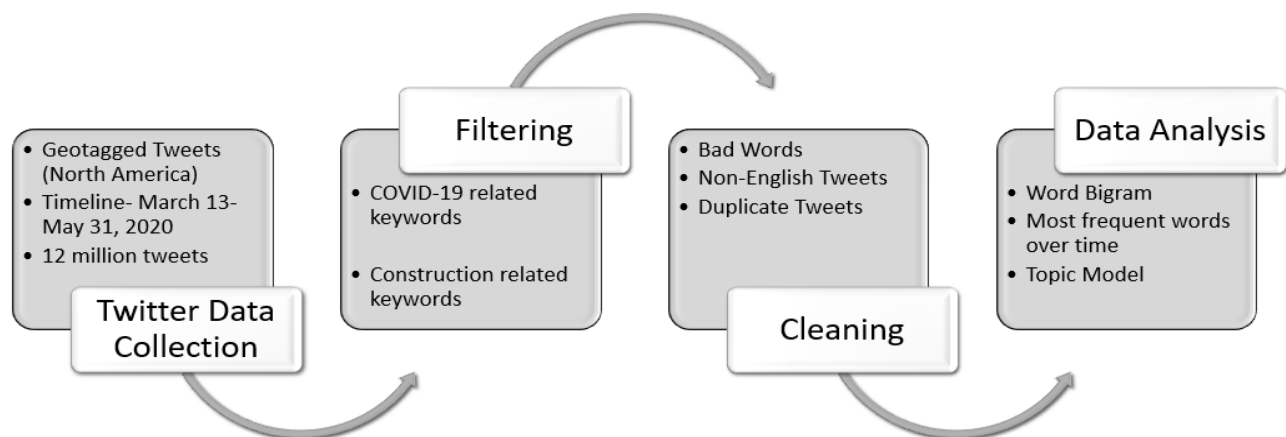


Figure 1. Flow chart of the study methodology

Besides, user name-name used by the user for a Twitter account, user location-user location, tweet date-time stamp at which tweet was created and clean tweets-tweet text except for bad words were also extracted. Different machine learning algorithms, such as word frequency count,

heatmaps, bigram, and topic modeling, were used on the clean data to achieve the corresponding results. A flow chart of the methodology followed in this study is shown in Figure 1. The discussions on the algorithms/ techniques used to obtain the results are described in the following sections.

DATA ANALYSIS

The word frequency model defines the most commonly used words which in turn helps to understand the user's communication patterns. We have deleted stop-words to boost our word frequency performance. The common terms that do not have any meaning and are incoherent to the model are considered to be stop-words. We used the 'nltk' library in Python to process such terms. The performance of the top 100 words used in the tweets is shown in Table 1. Words like “corona”, “covid”, “pandemic”, “coronavirus” were the most common keywords as most of the tweets are related to Corona. In addition to these terms, the outcome also includes construction-related topics such as labors, home, job, day, workers, house, patch, team, work, warehouse, productivity, staff, service, office, employees.

Table 1. Construction related most frequent words

<u>Rank</u>	<u>Frequent word</u>	<u>Frequency</u>	<u>Rank</u>	<u>Frequent word</u>	<u>Frequency</u>	<u>Rank</u>	<u>Frequent word</u>	<u>Frequency</u>
1	coronavirus	125795	19	need	6538	54	testing	3547
2	covid	122880	20	home	6357	66	bio	3145
3	pandemic	58798	21	job	6206	72	house	2957
4	corona	47554	22	day	6124	73	stayhome	2938
5	virus	15502	23	work	6088	83	safe	2731
6	people	14991	26	deaths	5401	86	working	2704
9	labor	10318	47	health	3975	91	workers	2646
13	time	8372	52	link	3563	98	looking	2515

The diagram in Figure 2 is a word network derived from the most frequent word bigrams of the datasets. In the network, the words serve as nodes and bigrams serve as connectors. A bigram is a pair of words that appear together in a text database. For example in this dataset, “covid” and “cases” have appeared together frequently, also words like “labor” and “job” have concurrently appeared frequently in the network. In the results that we obtained from word mapping using bigrams, we observe the dense network formed by the corona-related words and different topics have stemmed from central node words like “coronavirus”, “pandemic”, “positive”, etc. The longest chain of words propagated through the network is: *coronavirus-covid-crisis-testing-patients-cases-quarantine-news-deaths-response-global-spread-stayhome-outbreak-task-force-amid-pandemic-positive-confirmed-cases-fight*. The network link formed with words *opening-job-link-via-bio-click-apply* shows during pandemic many lost jobs and were looking for opportunities which are also echoed in the word link *recommend-anyone*. In the word network, we can see construction industry-related links such as *job-opening-labor-latest-team*, *package-handler-worker-warehouse* apart from the other topics with longer links.

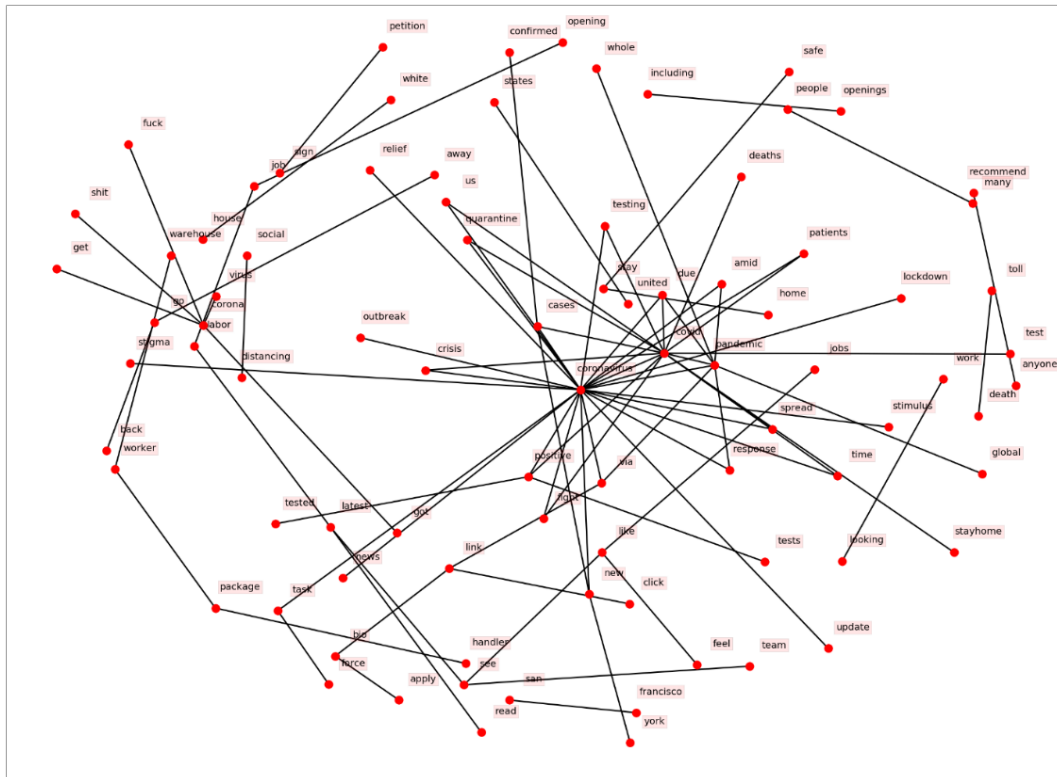


Figure 2. Word network using “bigrams”

A heatmap is a graphical data representation that uses a color-coding scheme to represent various values. Here, over a period of time, the heatmap graphically reflects the word frequency. That is, we have plotted the word frequency heatmap against the day, reflecting the word frequency of each day. The color intensity indicates the frequency of the word, the darker the color the frequency of that word is more. The heatmap in Figure 3 was recorded using the data in our dataset. From the heatmap we can see that the words Corona, coronavirus, covid, pandemic, people etc. were highly used throughout the period of three months. Apart from the words mentioned above we can also see that word cases, day, deaths, due, get, got, help, home, job, know, labor, need, new, one, stayhome, time, trump, us, via, virus, work workers were also used moderately over this period. Out of the list of words mentioned above the words day, due, help, labor, job, home, need, new time, work, workers are related to the construction industry. In the below heatmap we can see words such as business, casting, bill, distancing, done, hiring, looking, money, opening, etc. related to construction.

Topic model results

From the available dataset, we selected the tweets containing the construction-related terms such as construction, contractor, etc., and performed analysis on this subset of data. The results that we obtained on the subset are shown in Table 2. We obtained around 70 hidden topics in this sub dataset using Latent Dirichlet Allocation (LDA). LDA is a widely used topic modeling technique to extract topics from the textual data. Topic Coherence measures the score of a single topic by measuring the degree of semantic similarity between high-scoring words in the topic.

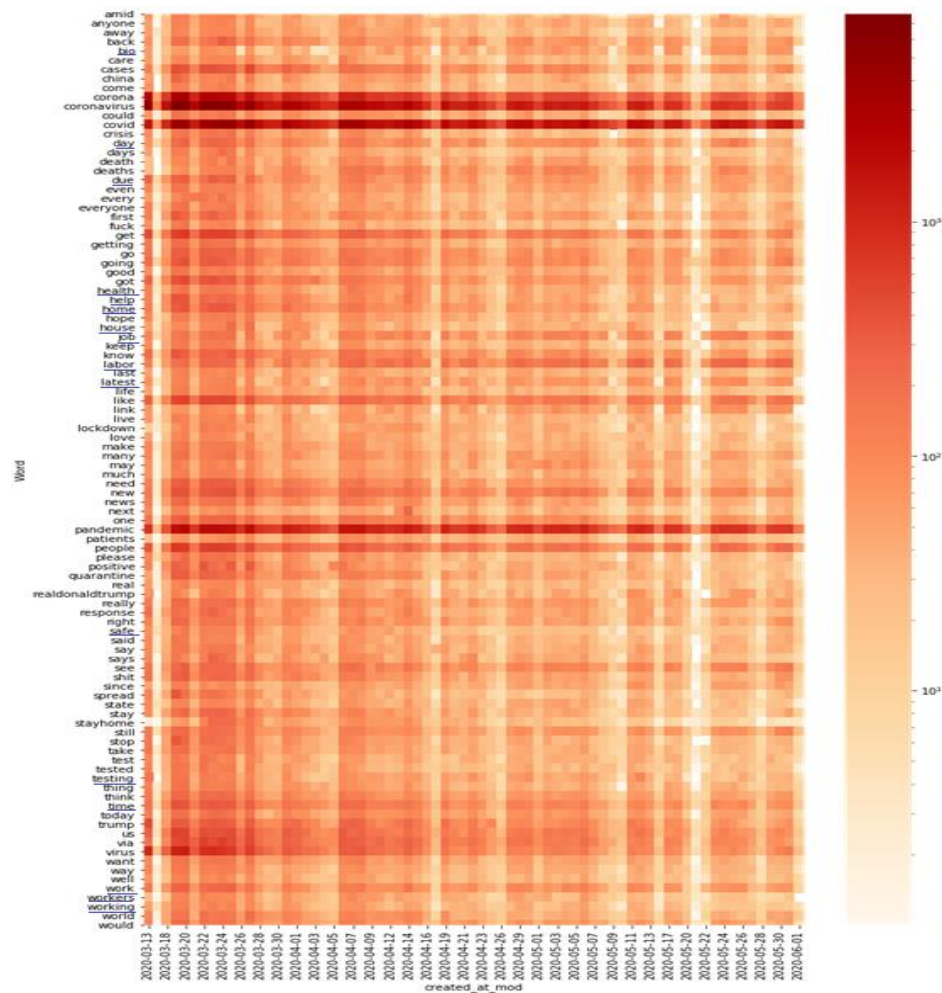


Figure 3. Heatmap for frequent words 1 to 100

The Figure 4 in following shows how 70 optimum topics are obtained. On the X axis, we have the number of topics and on the Y axis, we have the coherence value obtained for the corresponding number of topics. We can see that for 70 topics the coherence value is highest which is 0.566 and it decreases as the number of topics is increased from 70 to 80 i.e. beyond these seventy topics, the similar topics started to repeat. From each topic we have documented eight words in the order of higher to lower frequency of the word in the topic. From the results, we found that most of the topics were related to construction job opportunities, the effect of corona on the construction industry and labors. We found that most of the topics have the word labor in them. The first eight topics in the Table 2 are related to the job openings, hiring, and opportunities in the construction industry. Topic 21 represents a topic that shows unemployment during the pandemic. During the pandemic, many people lost their jobs and had to start the job search again, and as the market started opening slowly there were many job opportunities available and employers were posting job announcements on different sites including social media sites. The words apply, click, link, in, bio, latest, hiring, recommend, job, opportunity, etc. indicate the construction job opportunity tweets in the dataset. Along with the jobs in the construction industry, we can see that during pandemic many people wanted the home delivery of items for which employer needed warehouse package handler and a truck driver who would be willing to do a delivery of items.

Table 2. Topics related to the construction industry

Topic No	Topic tokens	Theme/ name of the topic
1	apply, click, labor, job, link, latest, hiring, opportunity	construction job opportunity
2	worker, warehouse, package, handler, labor, link, apply, hiring,	warehouse job openings
3	anyone, job, recommend, labor, see, link, bio, click	construction job recommendation
4	work, check, looking, labor, job, via, link, fl	job applications in florida
5	team, latest, opening, job, labor, including, via, ups	construction job hiring
6	read, latest, opening, job, hiring, via, labor, link, bio	construction job openings
7	like, open, see, job, currently, labor, link, bio,	job openings
8	teammate, latest, opening, see, job, including, labor, warehouse	latest job openings for labors
15	part, time, truck, driver, labor, trade, skilled, crane	driver jobs
9	casting, shit, perfect, director, city, ready, yeah, early	casting
10	productivity, today, home, thing, working, soon, amazing, profit	productivity of labors
11	day, labor, weekend, hard, coming, built, house, laboratory	labor work
12	patch, going, take, morning, unpaid, cabbage, auto, stupid	patch work time
16	patch, sour, kid, much, market, labor, thanks, report	patching job
17	know, patch, still, great, note, patching, big, point	patching job
18	really, good, labor, feel, like, business, park, land	work quality
30	labor, week, money, run, important, single, character, basic	labor weekly pay
31	better, labor, job, follow, title, appealing , product, opportunity	labor job opportunity
32	supervisor, labor, vote, cost, book, practice, human, hell	labor supervision
35	nc, need, country, labor, forklift, operator, straight, price	operator requirements
27	get, labor, cheap, migrant, exploit, amazon, walmart, highlight	migrant labor
33	slave, labor, ia, baby, long, healthcare, office, head	labor health
19	people, help, keep, labor, top, making, nike, environmental	people helping during pandemic
20	labor, hope, next, pandemic, world, emotional, solidarity, collaborating	people helping during pandemic
21	labor, service, department, coronavirus, unemployment, church, management, waste	corona effect on construction jobs
22	milestone, covid, food, social, community, school, together, collaboration,	corona effect on society
23	curing, state, value, corona, paid, virus, guess, labor	corona
24	demolition, man, ever, live, movement, energy, economy,	corona effect on economy
25	labor, free, black, probably, government, email, resource, zoom	corona effect
28	got, quarantine, labor, away, window, problem, industrial, discover	industrial problems in pandemic
34	entire, number, justice, disease, society, deserve, decision	disease outbreak
29	child, labor, law, getting, since, tax, god, preventing	labor law
47	stop, please, hospital, pay, remember, group, medium, inserter, labor	hospital pay for labors
26	delivery, call, labor, project, professional, specialist, contact	specialist employees
68	rate, excited, twitter, youtube, message, brain, realize, setting	app/ technologies

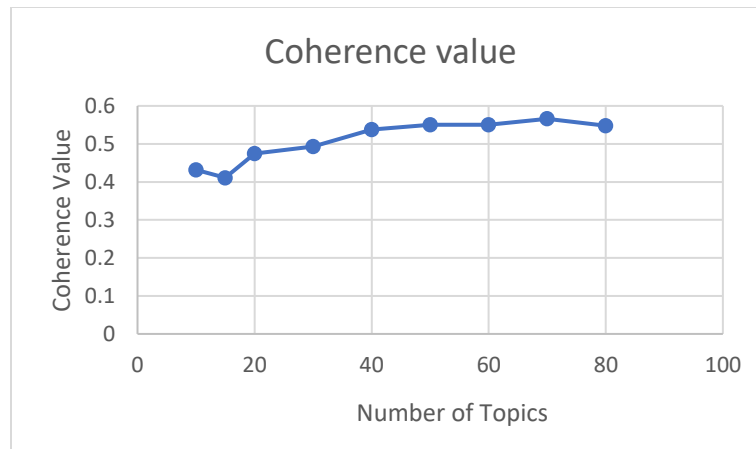


Figure 4. Coherence value for the number of topics

During the period when this data was collected, scientists were trying to find the vaccine for corona, and healthcare practitioners were using many different medications in order to cure corona. On topic 23 we can see the tweets about corona curing. As mentioned before, most of the topics contain the word “labor”. Due to its high frequency, we used the word “labor” for filtering the dataset for construction-related tweets. As we filtered the dataset for construction-related tweets and applied the topic modeling technique on it, we can see that many topics are labor specific such as the first twelve topics are about the labor jobs in construction or home delivery or warehouse jobs. Also, we have topics which are about labor health. For example, topic 33 discusses about labor healthcare and topic 47 has the word “hospital”, “pay”, and “labor” in it. If we trace these words back to the dataset, we can find many tweets about the health and safety of labor workers. Topic 29 has the words “labor”, “law”, and “child” as we have many tweets in the dataset which are about child labor as well as labor law. Also, during pandemic new labor laws have been added to the existing labor laws and some of these laws are about the paid sick leaves of the laborers during the corona pandemic. Topics 15, 9, 10, 11, 12, 16, 17, 18, 30, 31, 32, 35 are about the various services provided in the construction industry and were in run during the pandemic. Some of these topics are about the productivity of laborers, weekend work, patchwork jobs, and also the quality of work. In the dataset, we can also find some tweets about migrant laborers and shortage of migrant laborers creating inability to meet the demands, such as Topic 27 contains the word migrant labor.

From the Table 2, we can see many topics related to corona, such as topic 21 talks about the effect of corona on construction as well as many other industries, one of the harsh effects of corona was unemployment and many had to file unemployment in the labor department in order to receive some help or benefit from the government during the pandemic to keep going. We can see tweets related to the demolition of the economy due to the pandemic. Another effect of corona is that whenever any laborer or an employee is tested corona positive, he is supposed to be quarantined for a minimum of 2 weeks (Topic 28). During pandemic employees and as well as employers preferred to work from home and many applications such as zoom, google meet became important to conduct the meetings remotely. During the lockdown, people used many social media platforms such as Facebook, Twitter, and YouTube to spend most of their free time (Topic 25 and 68). During this period there was a dramatic increase in the usage of social media. As this data was collected from North America, we can find many cities and states in America along with Canada mentioned

in the tweets. These tweets are either about the construction jobs or work in these regions or about the covid situation or construction laborer conditions of these regions.

During the pandemic, most of the project managers focused on the work which can be executed from home such as documentation and design work. It was tough for employees to collaborate while working from home, but emerging technologies and applications such as zoom, Microsoft teams, WhatsApp chat groups helped employees to collaborate and carry out the tasks (Topic 25). As the schools were closed during the pandemic, it was a great opportunity for the construction industry to complete the major construction works in schools for which they usually need to wait till holidays. It is really inspiring how the construction industry turned this pandemic into an opportunity to finish pending works.

CONCLUSION

Twitter provides social media users a great forum not only to interact but also to express their thoughts to a wider audience, and thus Twitter provides an opportunity to research the reaction of people to crises. Twitter is often used to express thoughts or beliefs that different organizations use for critical correspondence relating to a person's health, economic or social well-being. Below are few key points which we can conclude from the analysis carried out:

- *The construction industry managed to keep up using various communication platforms. The industry adapted to the new ways of working such as using technologies for virtual connectivity to be functional during this pandemic.*
- *We observed that a small percentage of the employees in the construction industry worked from home but most of the employees who had to work on sites followed the rules such as wearing masks, maintaining social distancing, etc.*
- *We can see from the tweets that many employees from different industry including the construction industry lost their jobs during the pandemic, but the construction industry recovered from the pandemic and was looking for employees and hiring abundantly.*
- *During the period of lockdown, many contractors emphasized on the construction activities of schools as schools were closed. Usually, to perform major construction tasks in schools, contractors have to wait till the holidays.*
- *Through data gathering and converting data into functional formats, crisis informatics will benefit the overall preparedness and resilience during pandemic-like situations.*

Policymakers, corporations, and other industries would benefit from the findings as they would be able to take more efficient steps and be more resilient in present and future pandemics.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. IIS-2027360. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- Ahmed, M. A., Sadri, A. M., and Amini, M. H. (2020). "Data-driven Inferences of Agency-level Risk and Response Communication on COVID-19 through Social Media based Interactions." arXiv preprint arXiv:2008.03866.

- Ahmed, M. A., Sadri, A. M., Pradhananga, P., Elzomor, M., and Pradhananga, N. (2020). "Social Media Communication Patterns of Construction Industry in Major Disasters." *Construction Research Congress Proceedings-2020*.
- Bagheri, H., and Islam, M. J. (2017). "Sentiment analysis of twitter data." arXiv preprint arXiv:1711.10377.
- Broniatowski, D. A., Kerchner, D., Farooq, F., Huang, X., Jamison, A. M., Dredze, M., and Quinn, S. C. (2020). "The covid-19 social media infodemic reflects uncertainty and state-sponsored propaganda." arXiv preprint arXiv:2007.09682, 3(2).
- Camargo, K. R. D., Jr. (2020). "Trying to make sense out of chaos: science, politics and the COVID-19 pandemic." *Cadernos de saude publica*, 36, e00088120.
- Castillo, C., Mendoza, M., and Poblete, B. "Information credibility on twitter." *Proc., Proceedings of the 20th international conference on World wide web*, 675-684.
- CBO, C. B. O. (2020). "COVID-19 in Construction: Crucial Developments & Updates." <<https://www.constructionbusinessowner.com/covid19inconstruction>>.
- De Ceukelaire, W., and Bodini, C. (2020). "We need strong public health care to contain the global corona pandemic." *International Journal of Health Services*, 50(3), 276-277.
- Gros, C., Valenti, R., Schneider, L., Valenti, K., and Gros, D. (2020). "Containment efficiency and control strategies for the Corona pandemic costs. Cornell University."
- Haleem, A., Javaid, M., and Vaishya, R. (2020). "Effects of COVID 19 pandemic in daily life." *Current medicine research and practice*.
- Knight, H. A. (2020). "A Look at COVID-19 Impacts on the Construction Industry." <<https://www.hklaw.com/en/insights/publications/2020/05/a-look-at-covid19-impacts-on-the-construction-industry>>.
- Mojumder, M. N., Ahmed, M. A., and Sadri, A. M. (2021). "Identifying Ridesharing Risk, Response, and Challenges in the Emergence of Novel Coronavirus Using Interactions in Uber Drivers Forum." *Frontiers in Built Environment*, 7.
- Sadri, A. M., Hasan, S., Ukkusuri, S. V., and Cebrian, M. (2018). "Crisis communication patterns in social media during Hurricane Sandy." *Transportation research record*, 2672(1), 125-137.
- Sapul, M. S. C., Aung, T. H., and Jiamthapthaksin, R. "Trending topic discovery of Twitter Tweets using clustering and topic modeling algorithms." *Proc., 2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, IEEE, 1-6.
- Singh, L., Bansal, S., Bode, L., Budak, C., Chi, G., Kawintiranon, K., Padden, C., Vanarsdall, R., Vraga, E., and Wang, Y. (2020). "A first look at COVID-19 information and misinformation sharing on Twitter." arXiv preprint arXiv:2003.13907.
- Worldometer. (2020). "COVID-19 Coronavirus Pandemic data." <https://www.worldometers.info/coronavirus/country/us/>.