Journal of Information Systems and Technology Management - Jistem USP

Vol. 19, 2022, e202219008 ISSN online: 1807-1775

DOI: 10.4301/S1807-1775202219008 Original Article

ORGANIZATIONS AND EMPLOYEES SAY "I DO" TO WORK FROM HOME DURING THE PANDEMIC: A SENTIMENT ANALYSIS OF TWITTER

Surabhi Verma¹ https://orcid.org/0000-0001-7641-0637 **Vibhav Singh**² https://orcid.org/0000-0001-8810-2631

¹University of Southern Denmark, Odense, Denmark

²Narsee Monjee Institute of Management Studies University (NMIMS), Navi Mumbai, India

ABSTRACT

Social media platforms like Twitter have become an integral part of daily life. Twitter is emerging as a powerful digital platform for people to express their emotions and opinions on varied events and issues with the world. Working from home during the coronavirus lockdown is one phenomenon that has affected billions across the world, generating intense debates and polarizing global factions. This study aimed to analyze the sentiments of the Indian workforce about their work from home experience imposed during the coronavirus lockdown. To this end, a total 10,400 relevant English language tweets were collected from Twitter. Sentiment analysis and social network analysis were performed on the retrieved tweets. The results indicated that though the sentiments of the tweets were about consistently distributed among the three principal categories (positive, negative and neutral), there was a predominance of positivity towards working from home due to the coronavirus lockdown. Further, social network analysis of the tweets unraveled varied clusters indicating employee's sentiments towards the pandemic-induced work from home policy.

Keywords: Work from home, COVID-19, Twitter, Sentiment Analysis.

Manuscript first received: 2021-05-05. Manuscript accepted: 2021-11-10.

Address for correspondence:

Surabhi Verma, Department of Marketing & Management, University of Southern Denmark, Odense, Denmark.

Email: verma9016@gmail.com

Vibhav Singh, Narsee Monjee Institute of Management Studies University, Navi Mumbai. India, Email: drvibhav1108@gmail.com



INTRODUCTION

Around 3.4 billion people in 85 countries are grounded in their homes from March 2020, which translates into several millions of employees working from home (WfH) (Bouziri et al., 2020). Aiming to break the chain of the COVID-19 pandemic, Government of India advised all the organizations to implement a WfH policy for its employees (The Wire, 2020). In traditional HR literature, WfH is bracketed under the ambit of flexible work environment (Bloom, 2014). Aided by smart technology and motivated by the promise of sustainability, WfH has recently gained traction from companies due to varied reasons like economy of time, space and energy (Mas and Pallais, 2017). Employees find WfH advantageous as it permits them to spend more time with family, reduces the commutation time, encourages their creativity and helps avoid the workplace environment stress (Bloom et al., 2015). On the other hand, organizations believe that WfH aids them to hire the best in the market, negating geographical boundaries. For them, WfH helps attract high performing talent from competitors, cuts down the operational expenses, and acts as an enabler to the work-life balance philosophy (Fonner et al., 2012). However, despite numerous advantages associated with WfH, there have been critics of the policy as well. Many have argued that WfH usually gets translated into "actually not working" (Bloom, 2014). The traditional mental orientation of a manager is that there may be some loss of control, causing reduced work productivity of the employees (Raghuram and Wiesnfeld, 2005). Clearly, success of the WfH policy is contingent on the manager and the employee. A manager needs to be agile and demonstrate trustworthiness, whereas an employee needs to showcase a higher degree of dedication, discipline and self-control to ensure improved productivity while working from home (Bloom, 2014). Nevertheless, despite several companies having adopted flexible work arrangements for their employees, WfH remains a contested policy (Jones, 2017). One of the major reasons cited for the failure of the WfH policy is the lack of training imparted to employees by the organizations on how it may benefit them, how their productivity will be measured, and most critically, how to optimally function while working from home (Fonner et al., 2012).

Even as the debate over efficacy of the WfH policy was underway, the crisis precipitated by the coronavirus pandemic enforced a lockdown in nearly every country. Either voluntary or government-induced work from home policy had to be hastily imposed. Unfortunately, this created a knowledge vacuum, both at the employee and the organization level. It is important to understand that implementation of WfH under normal circumstances and as an emergency response to the unprecedented COVID-19 catastrophe are widely different premises (Baker, 2020). Also, while the pandemic-induced WfH policy aims to practice social distancing and avoid the virus spread, there are certain jobs that cannot be performed effectively from homes (Crandall & Longge 2005).

The present pandemic leading to the enforced WfH policy implementation has several caveats attached to it. First, telework was not anticipated by the employees and organizations and hence, both the stakeholders lacked readiness (Bouziri et al., 2020). Second, as the WfH came suddenly, social distancing led to forced self-isolation in several cases, with employees stranded away from home or left alone without family or friends. This may have a negative impact on their well-being (Bouziri et al., 2020). Third, most employees are faced with gripping anxiety regarding the impending economic crisis, and WfH has further augmented the psychosocial risks linked with telework (Bouziri et al., 2020). In this backdrop, the present study aims to investigate the Indian employees' opinion and sentiments towards work from home policy implementation during the pandemic, employing text data from Twitter. The main contribution of the study is to empirically analyze the employees' opinion towards the WfH policy in the unusual and anxiety-prone context of COVID-19.

Social media has permeated in the day-to-day functioning of people in the present times. It is a catalyst for building virtual connects amongst users through the sharing of opinions. People develop associations via tweets, retweets, comments, and likes (Kwak et al., 2010). While individuals who use social media often try to showcase their lives and thought processes through their profiles (Kwak et al., 2010), organizations generally use social media platforms to understand customer behavior and attempt to influence their buying behavior (Stieglitz and Xuan, 2013). In academia, social media has been investigated from multi-disciplinary standpoints like psychology, marketing, education, and sociology (Pang and Lee, 2008). Twitter is one of the most influential social media platforms with more than 500 million tweets being posted daily on an average (Statista, 2019). Around 330 million Twitter users are active per month, whereas the number of daily active users is 152 million; around 80% of the users use their cellular phones to tweet (Statista, 2019). These figures indicate that Twitter's penetration among global netizens is higher than other social media sites, and it provides deeper coverage of issues impacting people across geographies (Hughes et al., 2012). Therefore, in this study, we focused on the Twitter database for sentiment analysis regarding WfH during the coronavirus pandemic.

Though the WfH policy is a well-established HR intervention, its execution during a crisis like COVID-19 changes the dynamics and makes it rather complex. Gartner's report (2020) argued that 54% of Indian companies were ill-equipped to support WfH. It has been reported that companies are struggling to build and sustain their work culture during WfH (The Hindu, 2020). Microsoft CEO Satya Nadella proposed that WfH, in the longer run, may negatively impact employees' overall wellbeing. As there are contradicting perspectives on WfH during the crisis, it is important for academicians and HR practitioners to understand employees' perception about the same. The present study attempts to understand the sentiment of employees towards the WfH policy during the COVID-19 crisis. The study also attempts to identify the key themes underlying employees' sentiments.

The next section delves into the research methodology employed in the study, followed by an in-depth interpretation of the results. Moving further, the authors have critically deliberated upon the results in the discussion section, post which the potential implications and unavoidable limitations of the study have been discussed in the next section. Lastly, the insights derived from this study have been summarized in the conclusion section.

RESEARCH METHODOLOGY

Sentiment analysis is an emergent research field, which abstracts individuals' thoughts and behaviors from the text data through Natural Language Processing (NLP) methods (Öztürk and Ayvazb, 2018). Sentiment analysis is also referred to as opinion mining. In view of the exponential outreach of social media platforms, people have begun to share their thoughts and opinions through these platforms. Therefore, the opinions that users share on these platforms on any given issue can have substantial influence. In comparison to the traditional data mining techniques, sentiment analysis handles unstructured data (Oza and Naik, 2016). In recent times, data extracted from Twitter has become a popular avenue for academicians to conduct research (See works of Singh et al., 2020). Therefore, this study attempts to understand employees' sentiment regarding work from home (WfH) policy implementation during the COVID-19 pandemic using Twitter.

This study extracted tweets in the English language from Twitter and examined them in depth, using the R programming language and UCINET. Since India underwent the coronavirus lockdown



Verma and Singh

from 24th March 2020, the frequency of tweets on COVID-19 and WfH has increased exponentially on Twitter. The tweets are emerging as a valuable source of data yielding information about the challenges and benefits of working from home during the coronavirus crisis. In India, netizens generally use the English language for posting tweets, and therefore, the authors decided to analyze only English tweets about work from home during the coronavirus lockdown.

Data extraction

The data (all potential tweets) were searched and extracted using the TwitterR package from the R programming language. TwitterR gathered tweets between 24th March 2021 and 10th April 2021. Further, the retrieved tweets by TwitterR were saved in CSV files, which contained the information of tweets and the users posting them. Once all data related to the different hashtags were retrieved and saved in CSV files, the files were combined into a consolidated CSV file for further cleaning and preprocessing. The data file contains several elements other than tweets, including userID, dates, retweets, and global coordinates. Although these attributes are substantial for the analysis, not all tweets have complete information on all the aspects. Therefore, only those tweets that had no missing values in other attributes were considered for further analysis. Total 192,364 tweets were gathered with the hashtags related to coronavirus (e.g. #corona or #coronavirus or #covid-19 or #covid) and workfromhome (e.g. #workfromhome or #workathome or #remotework or #virtualwork or #telework or #distancework or #homeoffice).

Data cleaning and preprocessing

The retrieved dataset from TwitterR had several duplicate tweets as well as missing values in rest of the attributes. The Twitter application programming interface (API) is not a live stream, and it provides a set of tweets in a given timeframe. Therefore, the Twitter API mostly returns a pool of overlying tweets if the request to retrieve tweets is sent in short time intervals. According to Singh et al. (2019), retweets are usually not considered as a reflection of personal opinion for such studies. Thus, as this study aimed to ascertain the sentiment of individuals, the data cleaning process involved the deletion of retweets and their omission from further analysis. As a result, only 10,400 tweets, i.e., approximately 5% of the retrieved data, were considered for further analysis after the data cleaning step.

For the cleaning and preprocessing part, all the usernames, punctuations, numbers, URLs and blank spaces were discarded from the data, restricting it to meaningful English words for the analysis. In addition, all the uppercase letters in the data were changed into lowercase characters. After cleaning the text, retained tweets were divided into tokens. The tokens were searched in each sentence and token sentiment scores were ascertained. Further, these token sentiment scores were gathered to calculate the sentiment scores of the cleaned tweets.

Term frequencies calculation

Word Clouds is a tool for visualizing the summary of a large set of text data, and terms most frequently used are handy tools for tabularizing the most frequent keywords in the text data. This study generated a word cloud using the most frequently used terms to explore the enormous number of tweets related to work from home induced by the coronavirus pandemic (Figure 1). The word cloud



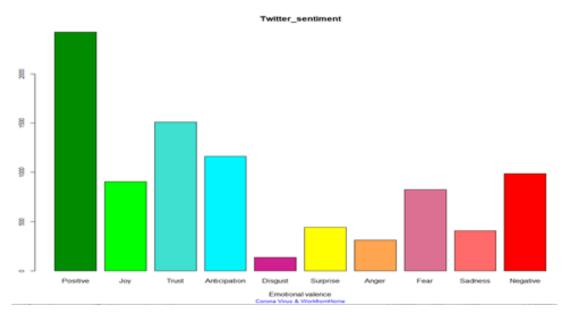


Figure 1. Sentiment analysis of tweets related to work from home and coronavirus

and most frequent terms were created by using the wordcloud and findfrequerms functions in the R programming language. In wordcloud, the size and location of words are determined in accordance with the frequencies of word usage in the retrieved tweets.

RESULTS

The sentiment scores of tweets were analyzed and grouped into different categories like positive, joy, trust, anticipation, disgust, surprise, anger, fear, sadness and negative.

Sentiment analysis of the Tweets

Syuzhet (a R language package) was used for the sentiment analysis of tweets. The Syuzhet package is especially designed for sentiment analysis and contains a sentiment lexicon to score the English words while calculating sentiments of the text data (Liu and Lei, 2018). This package classifies texts into ten different sentiment categories including positive, joy, trust, anticipation, disgust, surprise, anger, fear, sadness and negative. For each tweet in the dataset, an overall sentiment score was generated.

The sentiment scores of "Work from Home" Indian tweets were analyzed. Out of the 10,400 tweets analyzed, 1800 tweets were typified into the neutral category, i.e., anticipation and surprise. The highest number of tweets was categorized into the positive sentiment with 3000 tweets, in comparison to the negative sentiment that had only 1100 tweets. Therefore, positive sentiment was substantially higher than negative sentiment. Furthermore, nuanced understanding of sentiments could be demonstrated through positive sentiments like trust and joy having 1700 and 1000 tweets, respectively, whereas negative sentiments like disgust (200), anger (300), fear (800), and sadness (500) had fewer tweets. Figure 2 indicates the proportions of the identified sentiment groupings. Positive sentiment had the highest percentage of tweets (29%), followed by more positive sentiment tweets in the grouping of trust (16%) and joy (10%). 11% tweets fell under the negative sentiment



category. Negative sentiments like disgust were mirrored by 2% of the tweets; anger was displayed in 3% of the tweets, 5% of the tweets showed sadness, and 8% showed fear. Neutral sentiment grouping had 5% tweets for surprise and 13% for anticipation. Clearly, positive sentiments dominated over the negative and neutral sentiments.

Figure 3 indicates how varied terms under the category of positive and negative sentiments were frequented. Under the positive sentiment category, 'free' was the most commonly used term followed by 'productive', 'top', 'improve', and 'protect'. Words such as 'fun' and 'healthy' were the least used ones. Under the negative sentiment category, 'worry' was used most frequently, followed by 'restricted', 'tool', and 'symptoms'. The least frequented term under the negative sentiment category was 'break' followed by 'hard'.

The word count frequencies were analyzed for WfH Indian tweets with word clouds. Figure 4 below depicts the word cloud output for the same. The word "good" and "challenge" were the most dominant ones within the terms in WfH Indian tweet data. Therefore, they are highlighted bold in the cloud. It can be concluded that though the WfH policy in the current pandemic is welcomed by the employees, it is still bracketed as a challenge in terms of its implementation. The word cloud can be categorized into themes- (a) attitude towards WfH- worry, exciting, good, connect, help, learn, care (b) challenges emanating from WfH- productivity, security, risk-proofing, upskill, obedience. Figure 5 corroborates with the world cloud and cites words that have occurred 20 times or more in the WfH India tweets.

From the lens of social network perspective, the semantic relationship between dominant words is an important component of content analysis as conjoining such words in a text results into key meanings ascribed to the context (Park & Lee, 2009). Therefore, UCINET, a software package, was used to prepare the network visualization.

This study also delved into grouping of tweets into clusters. It is understandably difficult to categorize all the words into a cluster as some words can be grouped into multiple clusters in varied contexts; hence, the focus was on words that could be categorically connected with a cluster. Although it may be assumed that words in each cluster are appropriately representative of the clusters, there is

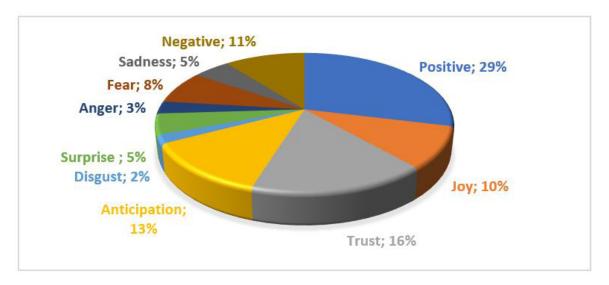


Figure 2. Percentage of tweets under each sentiment category

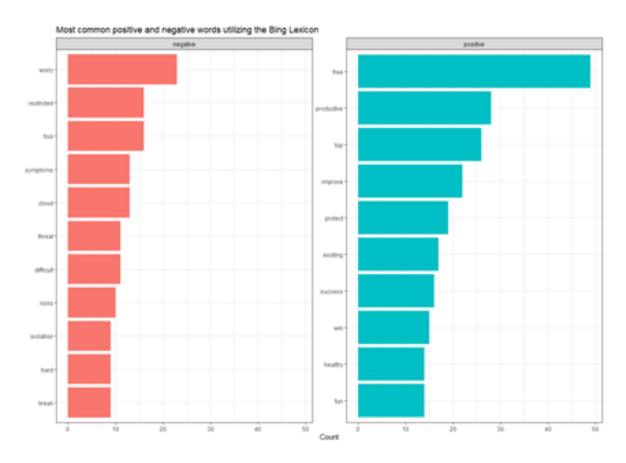


Figure 3. Most common Positive and Negative words



Figure 4. Word clouds of Work from Home tweets



a possibility that another word could be related with the clusters. The tweets were categorized into five main clusters (See Figure 6) - (a) Government response to Pandemic, (b) Employees' attitude towards WfH, (c) Structural support, (d) WfH impact on performance, (e) Leadership in crisis

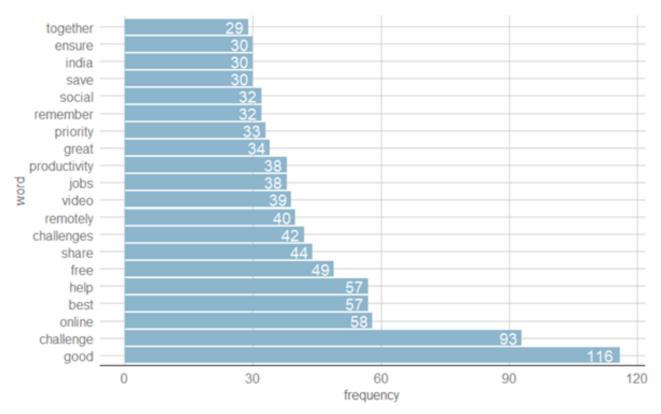


Figure 5. Overall frequency of words

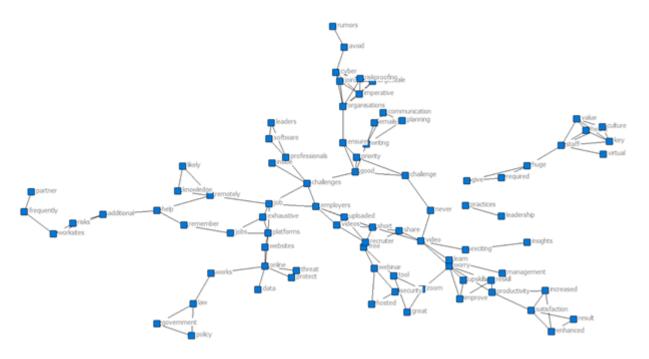


Figure 6. Network Visualization of Work from Home Tweets

DISCUSSION

In the present study, tweets of Indian employees about WfH policy execution during the coronavirus pandemic were analyzed and subsequently, sentiment analysis was carried out. The sentiment analysis of WfH tweets adds valuable insights by elucidating how employees are reacting to the WfH policy in these uncertain times, and the nature of challenges faced by organizations to ensure its effective implementation. In order to build a nuanced understanding of the phenomenon, the authors restricted the study to a developing economy like India. The overall analysis revealed that the highest number of tweets could be categorized under the positive sentiment. Within the positive sentiment, 'fun' was the most frequently used term, whereas under negative sentiment, 'worry' was the word used most often. The overall word cloud showed that 'good' and 'challenge' were the most highlighted words. The word cloud can be categorized into two themes- (a) Attitude towards WfH-worry, exciting, good, connect, help, learn, care (b) Challenges emanating from WfH- productivity, security, risk-proofing, upskill, obedience.

Furthermore, grouping of the tweets into categories resulted into five distinct clusters- (a) Government response to Pandemic: In order to control the number of cases in a densely populated country like India, the government decided to keep the country under a highly strict lockdown from 25th March 2020 onwards. This proactive decision by the government led to a large number of employees being confined and pressed to work from home. Rather strict penalties were levied on individuals and bodies for not adhering to the government guidelines. In order to ensure the safety of employees and reduce the threat of infection, companies and employees are appreciative of the WfH policy. In fact, many companies like Google have extended the WfH to the end of 2020. In this study, employees have used terms like 'Government', 'Law', 'Threat', 'Policy', and 'riskproofing'. This indicates that for most employees, the government acted vigorously at the policy-level to reduce the threat and risk associated with the spread of the virus. With the initial positive trends in India in restricting virus spread to specific areas, the lockdown was extended with a planned and phase-wise opening of the economy in the green zones. Also, the Indian government has provided a substantial financial stimulus in the form of a USD 270 billion dollars package for India Inc. and the marginalized sections of the society, which is 10% of the national GDP. (b) Employees' attitude towards WfH during pandemic: Hurton and Norman (2010) reported that organizational commitment of employees using the WfH policy was high in comparison to those who were not teleworking. However, another study indicated that in order to maintain positive relationships with co-workers, WfH employees invested more resources to be accessible for work all the time, leading to exhaustion (Bloom et al., 2015). In this study, it was observed that employees demonstrated a mixed attitude towards working from home. On one hand, terms like 'exciting', 'good', 'help' were used, indicating that employees believed WfH to be a positive step to ensure their well-being in the present crisis. However, usage of terms like 'worried', 'exhausted', and 'avoid' indicated the negative orientation towards WfH. Petriglieri (2020) argued that usage of platforms like Zoom and MicrosoftTeams requires higher degree of attention in comparison to face-to-face interaction. It is believed that during virtual meetings, an attendee has to work harder to decode the information and decipher verbal and non-verbal cues. Employees feel that they are under constant performance pressure, resulting into fatigue and exhaustion. (c) Structural support: one of the key aspects of successful implementation of the WfH policy leading to higher productivity is seamless adoption of technology. Studies show that virtual work success is contingent on whether the company provides appropriate tools and technologies to employees (Fonner et al., 2012). The parameters to assess the appropriateness of a technology include its ability to remove bottlenecks, reduce non-medical social



isolation, lead to quicker completion of tasks/projects, yield a higher degree of trustworthiness, and ensure alignment of technology with company's objectives (VernBurg et al., 2e013). In the present study, employees have used terms like 'Webinar', 'Zoom', and 'Videos', indicating that in order to maintain the essence of face-to-face communication, companies are adopting platforms with a video feature to maintain the connect. In addition, HR is using the lockdown period to manage the talent by ensuring employees' key skillset enhancement through online training and 'webinars' (d) WfH impact on performance: studies have shown mixed findings on the impact of WfH on productivity. Bloom et al. (2015) found that WfH increases the performance by 13%. However, even if there is an organizational support in the form of providing technical aid, HR support, manager's trust, and training and development, it positively impacts satisfaction but not productivity (Baker et al., 2007). Financial support offered by the employers significantly impacts employee's productivity (Baker et al., 2007). In the present study, terms like 'productivity', 'satisfaction', 'result', and 'upskill' were frequently mentioned by employees in their tweets, indicating that WfH policy implementation during the pandemic is impacting the organizational and employee-level outcomes. Therefore, while employees are taking the pandemic-induced WfH as an aberration, the general feeling is that this will have a lasting impact on their productivity and performance appraisal. (e) *Leadership in crisis*: much emphasis has been laid on developing the individual virtual worker, which is more the leader's responsibility than the worker's (Baker et al., 2007). Offstien et al., (2010) argued that success of WfH is more to do with the leadership than technology. It is important for the leaders to identify the enablers and derailers of WfH success (Naor et al., 2013). On the other hand, leaders have generally condemned the WfH policy on the account of productivity loss and changes in employee behavior with respect to their efficiency (Pepitone, 2013). However, in the present scenario, WfH is not an option that can be avoided but a necessity that has to be embraced. Therefore, the present scenario requires agility in leadership. In this study, employees have used terms like 'leaders', 'leadership', 'planning', and 'knowledge', indicating that the desired execution of WfH policy is not only contingent on the adoption of technology platforms by employees but also the role of leadership in this crisis.

Implications and Limitations

Given the backdrop of conflicting results regarding efficacy of the WfH policy, changes herein need to be investigated in detail. Therefore, based on the identified clusters, the authors would like to propose certain lines of inquiry that must be empirically tested by future studies. Studies in future need to investigate how structural support and leadership during crisis can impact employees' attitude towards WfH, thereby impacting the employee's productivity. Pandemic influence at the workplace in the near future has been significant, but it is unclear whether this influence will be long-lasting and cause strategic changes. Significantly, companies need to reassess the existing WfH policy to model it around 'WfH during crisis times'. Productivity metrics needs to realign to telework, wherein changes in working style due to environmental changes are factored in. Future studies need to investigate the factors that impact employees' functioning from home during crisis. Furthermore, studies need to unravel different types of adaptive strategies that anxiety-prone employees need to adopt to remain productive during these challenging times. HR managers and business leaders need to work around how they can customize jobs that are difficult to perform from home, for instance ground sales during crisis. For women employees, WfH has not been business as usual, with the strictest form of lockdown depleting their support system and disrupting their work-life balance.



Though the study contributes to building our understanding about a unique phenomenon such as the COVID-19 pandemic, there are certain limitations attached to it. First, the study did not keep varied locations into account while undertaking data collection by restricting the data to a specific geography. Being a timeframe analysis wherein data were collected during the strictest form of lockdown, there may be other reasons as well that could have contributed in shaping employees' outlook towards WfH.

CONCLUSION

In the present study, sentiment analysis of Twitter-retrieved data was carried out in the backdrop of the current pandemic to ascertain the general opinion of the Indian workforce towards WfH. Authors collected the desired tweets pertaining to WfH by people in India. This search resulted into a total of 10,400 tweets for further analysis post cleaning of duplicate and incomplete tweets. Overall sentiment analysis revealed that positive sentiment towards WfH was predominant in comparison to negativity, with significantly higher number of positive tweets. Furthermore, it was identified that tweets could be clustered into five themes indicating the employees' feelings about WfH- (a) Government response to Pandemic, (b) Employees' attitude towards WfH, (c) Structural support, (d) WfH impact on performance, (e) Leadership in crisis. In general, employees have built a positive outlook towards WfH during the pandemic, but then this policy is fraught with its own challenges in terms of overuse of technology, lack of agile leadership, anxiety, and possibility of social distancing leading to self-isolation.

REFERENCES

- Baker, M. G. (2020). Characterizing occupations that cannot work from home: a means to identify susceptible worker groups during the COVID-19 pandemic. MedRxiv.
- Bloom, N. (2014). To raise productivity, let more employees work from home. Harvard Business Review.
- Bloom, N., Liang, J., Roberts, J., & Ying, Z. J. (2015). Does working from home work? Evidence from a chinese experiment. Quarterly Journal of Economics. https://doi.org/10.1093/qje/qju032
- Bouziri, H., Smith, D. R., Descatha, A., Dab, W., & Jean, K. (2020). Working from home in the time of covid-19: how to best preserve occupational health? Occupational and Environmental Medicine.
- Crandall, W., & Gao, L. (2005). An update on telecommuting: Review and prospects for emerging issues. SAM Advanced Management Journal, 70(3), 30.
- Fonner, K. L., & Stache, L. C. (2012). All in a day's work, at home: Teleworkers'management of micro role transitions and the work-home boundary. New Technology, Work and Employment.
- Hughes, D. J., Rowe, M., Batey, M., & Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. Computers in Human Behavior.
- Jones, K. (2017). The Most Desirable Employee Beneifts. Harvard Business Review.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? Proceedings of the 19th International Conference on World Wide Web, WWW '10. https://doi.org/10.1145/1772690.1772751



- Liu, D., & Lei, L. (2018). The appeal to political sentiment: An analysis of Donald Trump's and Hillary Clinton's speech themes and discourse strategies in the 2016 US presidential election. Discourse, context & media, 25, 143-152.
- Mas, A., & Pallais, A. (2017). Valuing alternative work arrangements. American Economic Review.
- Öztürk, N., & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. Telematics and Informatics, 35(1), 136-147.
- Raghuram, S., & Wiesenfeld, B. (2004). Work-nonwork conflict and job stress among virtual workers. Human Resource Management. https://doi.org/10.1002/hrm.20019
- Shetty, D. (2020, May 18). Work-from-home is nice, but the office wants you back. The Hindu.
- Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. Social Network Analysis and Mining. https://doi.org/10.1007/s13278-012-0079-3