*Grenze International Journal of Engineering and Technology, Jul Issue*

# Twitter Sentiment Analysis of Public Opinions on Work from Home Methodology

Romit Bhalla[1], Prof. Sourabh Kumar Jain[2] and Lt. Prof. Praveen Kaushik[3]
[1-2]Gyan Ganga Institute of Technology & Sciences, Jabalpur, India
Email: romitbhalla.cs16@ggits.org, sourabhkumarjain@ggits.org
[3]M.A.N.I.T., Bhopal, India
Email: praveenkaushikmanit@gmail.com

***Abstract*— The 2020 world pandemic has brought into limelight an already existing and refurbished methodology of working known as 'Work from Home'. With preventive and curative treatments steadily reaching out most parts of the world now, the existence of Novel Coronavirus and its variants still cannot be denied. When more than 80% of the world was under lockdown and nearly a million people were at a risk of losing their lives, sectors like corporate, education, business and many others, on the recommendation of World Health Organization decided to allow their employees to work from remote locations where they can self-isolate themselves and prevent the spread on this infectious virus thereby some completely while others partially adhering to the work from home methodology. While being at home, people saw social media as a major platform to showcase their views and thinking over the ongoing and post-pandemic scenarios. The main aim of this study is to analyze public sentiments and emotions over Work from Home methodology based upon their views laid out on Twitter by using TextBlob, VADER and various Machine Learning Classification Algorithms to create a conclusive yet generalized picture of the presented positive, negative and neutral aspects of the viewpoints with the objective of identifying the most prominent one. Experimental evaluations show that people, in general, reflect 'positive to neutral' opinions towards the Work from Home methodology with Support Vector Classifier classifying the data into their respective sentiment classes with highest possible accuracy of 94.67%.**

***Index Terms*— Machine Learning, Sentiment Analysis, Opinion Mining, WordCloud, TextBlob, VADER.**

## I. INTRODUCTION

The year 2020 had been a crucial year over and in social media. The ups and downs that the pandemic brought are no secrets, but among these one thing remained consistent, it pushed more people to spend time online, expressing their views and opinions (positive or negative) over various issues they were facing and finding ways to engage and exchange these views with one another. One more reason for the continuous growth of social media and its popularity was the grand stage/platform it provided to keep track of the happenings around, friends, brands and family.

*Grenze ID: 01.GIJET.7.2.3*

*A. The Current Scenario*

With most of parts of the world shifting to the new normal and the initiation of the vaccination program in majority of the countries, most corporates employees are still following the Work from Home scheme to which they switched to during the pandemic lockdown though with slight variations. This scheme has been accepted with different outlooks by different people which they usually express on their social media handles. These beliefs and opinions reflect their personal analysis of the ongoing scenario, the benefits they received, the problems they faced and the limitations they encountered etc.

*B. The Statistics*

According to [1]., over 3.6 billion people use social media, with the expected projected number to increase to 4.41 billion in 2025. Reference [1] also stated that, internet users spend an average of 144 minutes on social media per day with Twitter emerging out as an important medium to stay updated with the latest circumstances around the world. Reference [2] reported 353 million monthly active Twitter users worldwide with the total number of tweets sent per day being 500 million at the rate of 9281 tweets per second. Due to this immense popularity and usage of Twitter, we will be using it as our main source of data for text analysis and natural language processing to get a clear picture of public opinions on Work from Home methodology in the pandemic and post-pandemic scenarios as a part of Sentiment Analysis process.

*C. Aim of this Research*

The main aim of this research is the analyze public attitude on Twitter towards Work from Home methodology, its impact on their lives, family, productivity, mental health etc. and reach to a non-exclusive conclusion by classifying their conceptions (positive, negative or neutral) using some common Machine Learning algorithms like Naïve Bayes, Support Vector Machine, Logistic Regression, Random Forest and Neural Network.

In Section II, we present the Literature Survey based upon the fields of applications of Sentiment Analysis. Section III describes the Proposed Methodology and Framework for Twitter Sentiment Analysis. In Section IV, we carry out an analysis of the results obtained thereby presenting Logical Conclusions and Future Scope of this research.

## II. Literature Survey

Reference [3] elucidates 'Sentiment Analysis' as *the field of examination that inspects people's opinions, sentiments, evaluations, appraisals, attitude and emotions towards units such as products, services, organizations, individuals, issues, events, topics and their attributes*. Some common synonyms for Sentiment Analysis are: *Opinion Mining, Sentiment Mining, Emotion Mining, Review Mining, Text Mining* etc. Reference [4] classifies Sentiment Analysis as *one of the primary use-cases of 'Natural Language Processing'(NLP)*. Reference [4] further defines (NLP) as *a branch of 'Artificial Intelligence'(AI) that studies how machines understand human language with the goal of building systems that can make sense of text and perform tasks like translation, grammar checking, or topic classification.* In order to automate these tasks, we require yet another component from under the umbrella of AI i.e. 'Machine Learning', designated by [4]. as *the process of applying algorithms that teach machines how to automatically learn and improve from experience without being explicitly programmed.*

Reference [5] performed 'Text-Mining' for deep analysis of knowledge in one's writings using an 'Ontology' based dictionary and 'WordNet' to classify text as positive, negative or neutral using NLTK Python library. Reference [6] used 'TextBlob', Python's text processing library along with MongoDB database and JSON in order to capture affinity relationships amongst users based on the content of their tweets. Reference [7] proposed a methodology to 'quantify sentiments' in Facebook by analyzing the users' emotions towards an issue based on the comments on the posts. The emotions (happy, sadness or emotionless) patterns were analyzed and presented using D3js tool.

Reference [8] compared the performances of various Machine Learning Algorithms for Prediction of Mental Health Problems among Higher Education students. Reference [9] studied the impact of Long Time Home Quarantine on Human Mental Behavior during COVID19 pandemic scenario using Machine Learning. Reference [10] investigated different Machine Learning Algorithms to determine Human Sentiment using Twitter data.

## III. PROPOSED METHODOLOGY AND FRAMEWORK

The stages of the proposed methodology for Twitter Sentiment Analysis of Public Opinions on Work from Home methodology are depicted in the Figure 1.

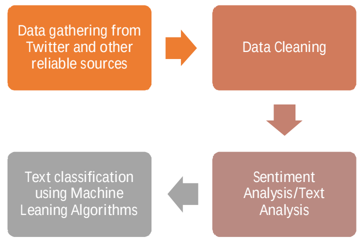


Figure 1. Proposed Methodology of Twitter Sentiment Analysis

### *A. Data Gathering from Twitter and other reliable sources*

According to the statistics provided by [11]., Twitter is the 6th ranked mobile app in 2021 so far, and the most 'digitally trusted' social media platform and the utmost reason for this success are the 'Twitter Hashtags'. Reference [12] define 'hashtag' (#) *as a symbol used to index keywords or topics on Twitter*. This function was created with the sole motive of allowing people to easily follow topics they are interested in. Reference [12] further added that Twitter allow its users to use the hashtag symbol (#) before a relevant keyword or phrase in their Tweet to *categorize those Tweets* and help them show more easily in Twitter searches. The primitive goal of our research is *to investigate public opinions and sentiments over the Work from Home methodology by eyeballing their tweets on the same.*

The data extraction process from Twitter is carried out by making use of 'Tweepy' Python Library. Reference [13] explicated *Tweepy as an open source Python package that gives you a very convenient way to access the Twitter API with Python*. Tweepy includes a set of classes and methods that represent Twitter's models and API endpoints as it transparently handles various implementation details such as: *Data encoding-decoding, HTTP requests, results pagination, OAuth authentication, Rate limits Streams etc*. By making use of 'Tweepy' Library we extract tweets relevant to *#workfromhome, #wfh, #workingfromhome and #homeoffice* in English Language, from a week ago as per Twitter's developer subscription norms.

As a secondary and summarized source of data, we choose the survey particulars carried out by [14].

### *B. Data Cleaning*

The data obtained in Stage 1 is 'noisy' and there is a very high probability that a subset of this data do not have any major contribution in the further stages of the Sentiment Analysis Process, and hence is required to be cleaned up. For example: a general tweet along with text also comprises of *URLs(links), images, punctuations, special characters, acronyms, emojis and other hashtags* which can be categorized as 'noise' for not adding any value to the tweet and should be removed, replaced or handled using an 'appropriate algorithm'. The steps involved in the data cleaning process are described below:

*1. Tweets' Segregation:* The tweets are manually segregated (although not precisely) based upon the subjects listed by [14].

*2. Conversion of text to Lower Case and splitting up of compound words:* This step is necessary in order to bring uniformity in text and avoid any future ambiguity in Sentiment Analysis and Text Classification phases.

*3. Removal of noise from data using Regular Expression Library:* This includes removal of URLs, images, punctuations, hashtags, re-tweet identifiers and emojis.

*4. Tokenization and Stop words removal*

*a. Tokenization:* The process of chopping off a large piece of text(corpus) into smallest constituent meaningful units is termed as Tokenization. Here, we tokenized the collected text into constituent words called as 'tokens' using 'Word Tokenizer Library'.

*b. Stop Words Removal:* The process of removal of irrelevant, fillers and words which do not contribute any significant meaning to the Sentiment Analysis process is called as Stop Words Removal.

*5. Stemming and Lemmatization*

*a. Stemming:* The process of extracting the 'root' or 'stem' word, which adds the same meaning and significance to the text as the given form of word is called a Stemming.

*b. Lemmatization:* The process of Lemmatization is very much similar to that of Stemming, however the end result may be a synonym to the original word adding the same sentiment and emotional quotient to the text.

*C. Sentiment/Text Analysis*

Reference [15] define 'Text Analysis/Sentiment Analysis' *as the process of measuring out the polarity of a piece of text with the goal of bringing out the hidden emotions, sentiments and opinion polarity behind the clean and pre-processed form of the text*. The polarity of a text can be positive, negative or neutral and can convey emotions like love, joy, surprise, anger, fear, sadness etc. We sub-divided the Sentiment Analysis stage further into 4 phases as shown in Figure 2.

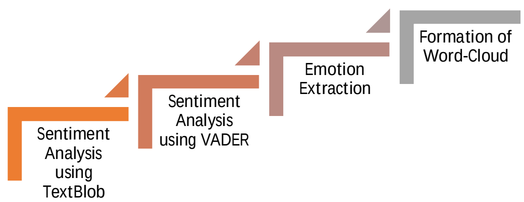


Figure 2. Phases of Sentiment Analysis Process

*1. Sentiment Analysis using TextBlob:* The cleaned out dataset is tested for 'sentiment polarity' using the 'TextBlob' Python Library. Reference [16] define 'TextBlob' as *a Python(2 and 3) library for processing textual data by providing a simple API for dealing with common natural language processing(NLP) tasks such as POS tagging, noun phrase extraction, sentiment analysis, classification, translation and more*. Reference [16] further mention that the 'sentiment property' returns a named tuple of the form Sentiment (polarity, subjectivity). The 'polarity score' is a float within a range [-1.0, 1.0] where the value -1.0 of the polarity score indicates 'extremely negative sentiment', 0 indicates 'neutrality' while +1.0 indicate 'extremely positive sentiment'. The 'subjectivity score' is a float within the range [0.0, 1.0] where 0.0 indicates extremely 'objective' and 1.0 indicates extremely 'subjective' form of conveying the message.

*2. Sentiment Analysis using VADER:* For comparative analysis, the cleaned dataset is tested for sentiments using VADER (Valence Aware Dictionary and sEntiment Reasoner), *a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media, designed and developed under [17]*. The 'compound score' of VADER is a float which when greater than 0.05 indicates 'positive sentiment', when less than -0.05 indicates 'negative sentiment' and when in range {-0.05 and +0.05} indicates 'neutral sentiment' based on the summation and normalization of valence score of each lexicon.

*3. Emotion Extraction:* To extract out the various emotions constituting public opinions towards the 'Work from Home' methodology and their respective quotients we use a dictionary provided by [18]. Each word in the dictionary is attached with a particular 'emotion'. The existence of each word in the tweet is checked against the dictionary and its corresponding emotion is noted and accumulated. Lastly, the count of each emotion obtained from the experiment is converted into corresponding percentage and is plotted on a suitable graph as shown in Figure 3.

*4. Formation of Word-Cloud:* A Word-Cloud is *a pictorial representation of words with higher significance, emphasis, importance, highlights and frequency than others*. A Word-Cloud of most frequently used words present in tweets in relevance to *#wfh, #workfromhome, #workingfromhome and #homeoffice* is depicted in Figure 4.

*D. Text Classification using Machine Learning Algorithms*

Reference [19] define 'Text Classification' as *a task of delegating a set of pre-defined categories to free-text*. 'Machine Learning Based Text Classification Systems' specifically, make classification based upon past observations. These systems make use of 'pre-tagged' texts as examples (training dataset) to classify new and unseen data (testing dataset) into various pre-defined classes. The various Machine Learning Algorithms used in our research along with their implementation details are shown in Table I.

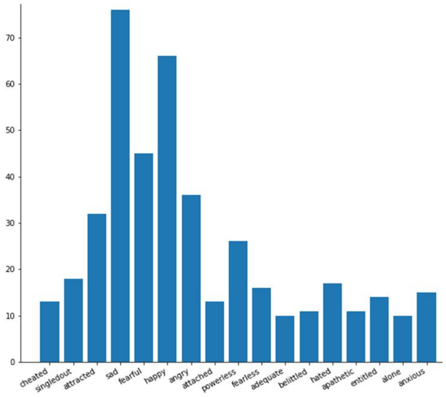


Figure 3. Experimental Results showing the quotient of each emotion extracted out of the tweets

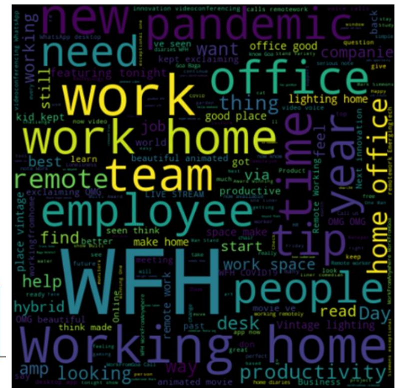


Figure 4. Word-Cloud representing the most frequently used words in tweets relevant to Work from Home methodology

TABLE I. THE VARIOUS MACHINE LEARNING CLASSIFIERS USED IN THIS RESEARCH ALONG WITH THEIR IMPLEMENTATION DETAILS

| **Machine Learning Classifier** | **Implementation details** |
|---|---|
| Naïve Bayes Classifier | Train Test ratio 90:10 |
| Logistic Regression Classifier | L2 Regularization |
| Support Vector Classifier | Train Test ratio 70:30 |
| Random Forest Classifier | Maximum Depth = 2 |
| Neural Network Classifier | 5 hidden layers with 2 neurons each |

## IV. CONCLUSIONS AND FUTURE SCOPE

### *A. Experimental Results and Discussions*

Table II shows the experimental results of Subjective Sentiment Analysis of tweets on Work from Home methodology using 'TextBlob' and 'VADER' respectively.

TABLE II. EXPERIMENTAL RESULTS OF SUBJECTIVE SENTIMENT ANALYSIS OF TWEETS ON WORK FROM HOME METHODOLOGY USING 'TEXTBLOB' AND 'VADER'

| Effects of WFH on | Percentage of Tweets | | | | | |
|---|---|---|---|---|---|---|
| ↓ | *Positive* | | *Negative* | | *Neutral* | |
| | *TEXTBLOB* | *VADER* | *TEXTBLOB* | *VADER* | *TEXTBLOB* | *VADER* |
| *Adapting to the New-Normal* | 25 | 0 | 25 | 25 | 50 | 75 |
| *Collaboration skills* | 55.55 | 11.11 | 11.11 | 22.22 | 33.33 | 66.66 |
| *Discovering new hobbies* | 71.42 | 57.14 | 14.28 | 0 | 14.28 | 42.85 |
| *Distractions* | 33.33 | 22.22 | 22.22 | 55.55 | 44.44 | 22.22 |
| *Relationship with family members* | 80 | 60 | 0 | 0 | 20 | 40 |
| *Favorability* | 40 | 80 | 0 | 0 | 60 | 20 |
| *Independency* | 66.66 | 66.66 | 0 | 0 | 33.33 | 33.33 |
| *Reduced interaction with society* | 40 | 60 | 40 | 20 | 20 | 20 |
| *Mental Health* | 25 | 25 | 50 | 50 | 25 | 25 |
| *Practicality in life* | 40 | 60 | 40 | 20 | 20 | 20 |
| *Priorities* | 20 | 40 | 40 | 40 | 40 | 20 |
| *Productivity* | 71.42 | 42.85 | 14.28 | 14.28 | 14.28 | 42.85 |
| *Time for self* | 75 | 0 | 0 | 0 | 25 | 100 |
| *Work Set-up* | 60 | 60 | 20 | 20 | 20 | 20 |
| *Suitability* | 83.33 | 66.66 | 0 | 0 | 16.66 | 33.33 |
| *Equipped work setup* | 0 | 16.66 | 16.66 | 0 | 83.33 | 83.33 |
| *Work life balance* | 50 | 33.33 | 0 | 0 | 50 | 66.66 |

The graph in Figure 3 shows that '*sad*' has been the '*most dominant*' emotion of the people while tweeting in relevance to *#wfh, #workfromhome, #homeoffice and #workingfromhome.* The reason for the same can be closely related to the statistics of the rising number of COVID19 cases yet again in most parts of the world, indicating the emergence of its 'second wave'. However, this can also be explained through the *subjective analysis of tweets* in Table II which show that the discussed working methodology has a '*negative impact*' on the *collaboration skills and mental health* of the people due to the *distractions* (inconsistent internet connection, power outage etc.) and exhaustion they have to face during their working hours. The results also show that people do face problems in *sorting out their priorities* while working from home due to lack of strict boundaries between working and non-working hours.

A comparison between the results of Twitter Sentiment Analysis of Public Opinions on WORK from HOME methodology using 'TextBlob' and 'VADER' is shown in Figure 5 which depicts that whether analyzed using 'TextBlob' or 'VADER' Sentiment Analyzer, people have shown '*positive response*' towards 'Work from Home' methodology and its repercussions, followed by '*neutral*' and '*negative*' responses respectively.

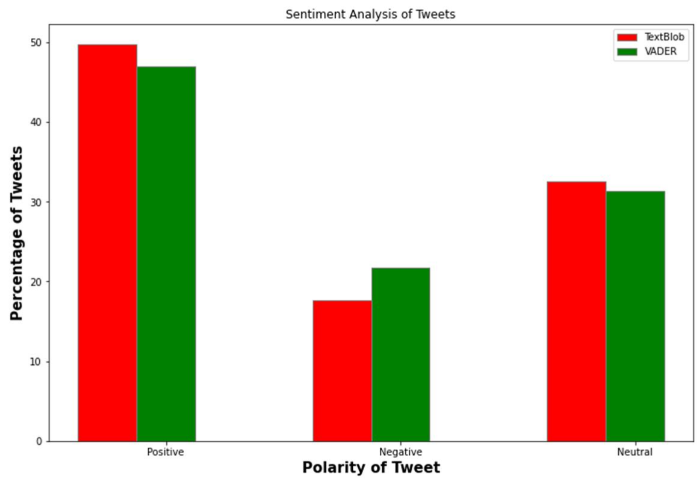


Figure 5. Comparison of results of Sentiment Analysis performed using TextBlob and VADER

The other key findings of the research reveal that employees *feel more independent* while working from home, get *more time to spend on themselves discovering new hobbies* and talents of their own. Also, the discussed working methodology has significantly *improved their productivity, practicality*, and *relationship with their family members* as they now get to spend more quality time with them. Surprisingly, the research reveal that the people have shown *'positive response'* over the *reduction on the amount of interaction* they have with the society which may be indirectly related to the mental exhaustion they face while working from home.

However, people have shown '*neutral behavior*' towards the lack of stagnant and well-equipped work set-up, their ability to adapt to the changes in the new-normal and maintain a constant work-life balance.

The word-cloud in Figure 4 reveals that words like *work, home, office, WFH, working home, team* etc. have been the most frequently used ones in the tweets relevant to work from home methodology. These words necessarily show the outlook of the employees towards their companies.

The graph in Figure 6 shows the respective prediction accuracy of each Machine Learning based Text Classifier used in our research.

The results plotted in the graph reveal that the *Support Vector Machine Classifier(accuracy=94.67%)* classified tweets most accurately, followed by *Neural Network Classifier(accuracy=94%), Naïve Bayes Classifier(accuracy=90.67%), Logistic Regression(accuracy=90%)* and *Random Forest(accuracy=85.33%)*.

### *A. Conclusions*

After scrutinizing all the results of the processing applied on the information source, it is safe to conclude that people, in general, feel *'contented and comfortable'* working from home, even during the post-pandemic scenario and the credit for this must go the technological advancement of the infrastructure and the internet.

### *B. Future Scope*

We expect that our findings can prove out to be useful for companies and institutions of any field for the benefits of the society. The experimental results can be used by various government and non-government organizations to bring about amendments and reforms in the working methodologies in order to ensure minimal financial loss to the organizations as they make a shift from pandemic to post-pandemic scenario. The generated knowledge is expected to have a wider reach for improvements and use in future works as Text Analysis holds bigger and brighter opportunities with social media being a reliable and convenient source of exploitable information.

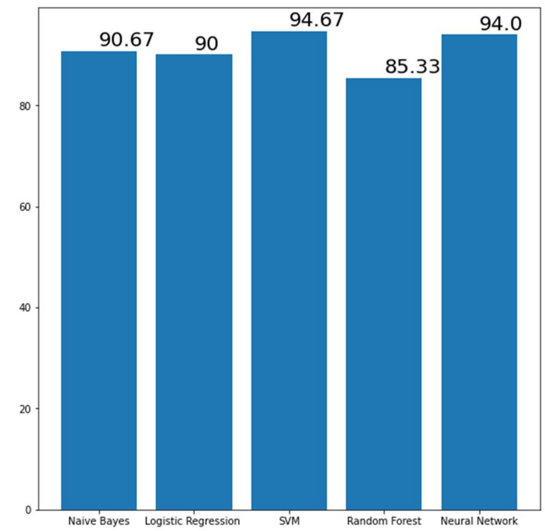


Figure 6. Comparison of prediction accuracies of various Text Classification Machine Learning Algorithms

ACKNOWLEDGMENTS

We would like to present our sincere thanks to the faculties and staff members of Gyan Ganga Institute of Technology & Sciences, Jabalpur, India for their necessary support and recommendations.

REFERENCES

[1] J. Chen, "36 Essential social media marketing statistics to know for 2021," Sprout Social, 2021. [Online]. Available: https://sproutsocial.com/insights/social-media-statistics/. [Accessed 2021].

[2] J. Bagadiya, "367 Social Media Statistics You Must Know In 2021," Social Pilot, 2021. [Online]. Available: https://www.socialpilot.co/blog/social-media-statistics#tw-usage-stats.

[3] Liu and Bing, "Sentiment Analysis and Opinion Mining," [Online]. Available: https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf.

[4] I. Roldos, "NLP, AI, and Machine Learning: What’s The Difference?," [Online]. Available: https://monkeylearn.com/blog/nlp-ai/.

[5] E. K. E. E. M. E. and K. A. E. , "Deep analysis o/knowledge in one's writings," in *2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS'15)*.

[6] A. R. D. F. and I. S. , "AffinityFinder: A System for Deriving Hidden Affinity Relationships on Twitter Utilizing Sentiment Analysis," in *2016 4th International Conference on Future Internet of Things and Cloud Workshops*.

[7] R. M. and S. D. , "Product response analytics in Facebook," in *International Conference on Intelligent Computing and Control Systems ICICCS 2017*.

[8] Nor Safika Mohd Shafiee, Sofianita Mutalib. " Prediction of Mental Health Problems among Higher Education Student Using Machine Learning ", *International Journal of Education and Management Engineering (IJEME)*, Vol.10, No.6, pp.1-9, 2020. DOI: 10.5815/ijeme.

[9] Imrus Salehin, Sadia Tamim Dip, Iftakhar Mohammad Talha, Ibrahim Rayhan, Kanij Fatema Nammi, "Impact on Human Mental Behavior after Pass through a Long Time Home Quarantine Using Machine Learning", *International Journal of Education and Management Engineering (IJEME)*, Vol.11, No.1, pp. 41-50, 2021. DOI: 10.5815/ijeme.

[10] Golam Mostafa, Ikhtiar Ahmed, Masum Shah Junayed, "Investigation of Different Machine Learning Algorithms to Determine Human Sentiment Using Twitter Data", *International Journal of Information Technology and Computer Science(IJITCS)*, Vol.13, No.2, pp.38-48, 2021. DOI: 10.5815/ijitcs.

[11] C. Newberry, "36 Twitter Stats All Marketers Need to Know in 2021," 2021. [Online]. Available: https://blog.hootsuite.com/twitter-statistics/.

[12] T. D. Team, "How to use hashtags," [Online]. Available: https://help.twitter.com/en/using-twitter/how-to-use-hashtags.

[13] M. Garcia, "How to Make a Twitter Bot in Python With Tweepy," [Online]. Available: https://realpython.com/twitter-bot-python-tweepy/#what-is-tweepy.

[14] Agngarayngay and A. Christian, "Perceived Effects of Work from Home setup among private employees in NCR," [Online]. Available: https://www.academia.edu/45144968/Perceived_Effects_of_Work_From_Home_Setup_Among_Private_Employees_in_National_Capital_Region_NCR_.

[15] R. Bhalla, P. S. K. Jain and P. P. Kaushik, "A Review Of Various Sentiment Analysis Techniques, Methodologies And Their Applications," in *11th International Joint Conference on Advances in Engineering and Technology*, 2021.
[16] Loria and Steven, "TextBlob," 2020. [Online]. Available: https://textblob.readthedocs.io/en/dev/..
[17] Hutto, C.J. & Gilbert and E.E., "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," in *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.*,
[18] A. Bhatt, "Sentiment Analysis with and without NLTK," [Online]. Available: https://github.com/attreyabhatt/Sentiment-Analysis.
[19] "Text Classification," Monkey Learn, [Online]. Available: https://monkeylearn.com/text-classification/.
[20] V. I. L. Rodriguez, "Home Office due to COVID-19 - Twitter Data Analysis".

## SOURCE CODE

https://drive.google.com/file/d/16NgTiKDRbH5Gdyptz0c57M273_hVZG0q/view?usp=sharing