Evaluating Occupant Feedback on Indoor Air Quality Perception During Covid Stay-at-home Using Social Media Data: A Nationwide Study in the U.S.

Mehdi Ashayeri (mehdi.ashayeri@siu.edu)
Southern Illinois University Carondale https://orcid.org/0000-0002-8364-9360
Soroush Piri
Southern Illinois University Carondale
Narjes Abbasabadi
University of Washington

Article

Keywords: Stay-at-home, Occupant Feedback, Indoor Air Quality Perception, Twitter, Sentiment Analysis (SA)

Posted Date: February 2nd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2481649/v1

Additional Declarations: There is **NO** Competing Interest.

Evaluating Occupant Feedback on Indoor Air Quality Perception During Covid Stay-at-home Using Social Media Data: A Nationwide Study in the U.S.

ABSTRACT

Despite challenges, the COVID stay-at-home has provided opportunities to reveal unforeseen complexities within built environments, particularly from indoor health lenses, that had not been sought before the pandemic. This research aimed to evaluate occupants' feedback on impacts of the stay-at-home on the indoor air quality (IAQ) perception in buildings nationwide in the U.S. during the first year of the pandemic (2019) and compare it with the baseline (2019). We used geo-tagged big textual data obtained from Twitter platform and developed Natural Language Processing (NLP) models based on Sentiment Analysis (SA) approach to compute the occupant feedback for the two consecutive years. We built the SA models through developing a six-step workflow, including data acquisition; data cleaning; text tokenization; analysis; accuracy evaluation; and data visualization and mapping. We used the *QDAP* dictionary and *nrc* lexicon to develop the SA models. We also developed automation scripts to improve the simulation performance. Results illustrate that occupants' complaints on IAQ increased during 2020 compared with the baseline (2019). Findings further suggest that occupants with less access to operative Heating, Cooling, and Air Conditioning (HVAC) systems posted more dissatisfied feelings on Twitter. This research aids decision-makers to better understand the impacts of lockdown on occupants' health experience to rethink the design and operation of buildings in a way to promote human-centered design and improving building resilience against future pandemics.

KEYWORDS

Stay-at-home, Occupant Feedback, Indoor Air Quality Perception, Twitter, Sentiment Analysis (SA)

1.0. INTRODUCTION

During the COVID-19 pandemic, the stay at home became a routine which helped prevent exposure to SARS-Cov2 viruses and mitigate the transmission rate [1]. According to the American Time Use Survey [2], the remote work during Covid-19 pandemic increased from 24% to 38% in 2020 compared with the baseline (2019). Due to the lockdown, in the U.S., people spent most of their time indoors: 90% in urban areas, 86% in the suburbs, and 82% in rural areas from March to April 2020 [3]. Accordingly, occupants started to complain on their buildings environments that they were not aware of them before the pandemic because the related challenge had not been raised in such scale [4]. Indoor Air Quality (IAQ), that directly impact the health and well-being of building occupants [5], is among those building characteristics that occupants have largely expressed complaints on them during the lockdown. IAQ measures the quality of indoor air in a building which affects the comfort and well-being of occupants. It is evident that exposure to indoor air pollution can negatively impact occupants' health (e.g., respiratory and cardiovascular diseases [6–8] [9], cognitive malfunction [10] and more).

In addition, IAQ boosts virus transmission rates in buildings. Most recently, findings (e.g., [11]) indicate that the use of cleaning supplies in buildings along with the poor ventilation condition during the Covid-19 pandemic led to higher indoor pollution concentrations beyond the standard level [12]. The impacts of indoor pollution concentrations on occupants' health rose throughout the quarantine due to changes in outdoor concentrations level elevated before the pandemic and exposure to outdoor-origin indoor PM2.5 concentrations in buildings [13]. Among indoor pollutants, studies indicate that exposures to particulate matters (e.g., PM2.5) associated with higher mortality rates since PM2.5 particles can help carry the SARS-Cov2 viruses [14]. It further increases the severity of Covid-19 symptoms [15]. Therefore, indoor spaces needed to be properly ventilation, particularly with fresh air, to lower the virus spread rate and the risk of infection [16] for future similar scenarios.

During the lockdown, groups of people were dissatisfied about the lack of performative Heating, Cooling, and Air Conditioning (HVAC) systems in their buildings [17], because those buildings have not been designed for such long-term stay-at-home. HVAC systems are generally designed to maintain the temperature, humidity, and contaminant levels within the standard range, therefore, modifying them to function under longer periods during the pandemic involved in lower efficiency and higher operation and maintenance costs [18]. Installing filters within the air-intake pathways can limit the penetration of outdoor particles toward indoor [19]. The high-efficiency particulate air (HEPA) purifiers could facilitate providing clean air environment for occupants with acceptable HVAC-operation performance [20]; however, they can partially protect the virus transport toward indoors. Because, current HEPA filters can only remove particulates of lesser than 0.3μm in aerodynamic dimeter, while SARS-Cov-2 viruses are equal or lesser than ~0.1μm size [21].

Moreover, studies confirm that there exists a direct correlation between higher temperature levels and IAQ dissatisfaction during the pandemic (e.g., [22,23]). An experimental study by [24] emphasize that a high temperature in indoor spaces increases VOC (volatile organic compounds) emissions, mostly driven by chemically-produced building materials, which result in poor IAQ, suggesting that occupant feedback on IAQ occur more in the summer time, followed by the spring and winter. Negative feedback drops in winter due to the temperature decline [25,26]. Guo et al. [27] obtain data from the USGBC database and explore occupants' feedback on IEQ factors of buildings certified by Leadership in Energy and Environmental

Design (LEED) compared with the non-certified buildings, using text mining approaches. They find occupants' feedback almost with the same polarity score (negative/positive) for both categories.

Most recently, occupant feedback collected from online big data platforms has open up an avenue for indoor health-related studies. For example, research by [41] collect data from online reviews of hotels on reservation websites and apply text mining techniques to capture occupant complaints on IEQ level in hotel rooms and find direct relationships between IAQ and seasonal weather variations as well as differences between IAQ satisfaction and climate zones. And study by [29] extract big textual data from Airbnb reviews upon visitors' feedback on IEQ experiment during the their stay and apply text-mining approaches to measure the sentiment score. Similarly, study by [28] implement spatiotemporal text mining approaches to capture and evaluate occupants' feedback on IEQ dissatisfaction.

Lately, advances in data-driven techniques have helped develop more reliable models for the human-centric design decision-makings. Text mining as a data-driven approach is a technology that uses Natural Language Processing (NLP) frameworks allowing machines to understand the human language, primarily through transforming unstructured text into normalized and structured dataset suitable for analysis [30]. Started in the 1950s [31] at the intersection of linguistics and AI [32], NLP was initially distinguished from text information retrieval (IR), which uses highly scalable statistics-based techniques to effectively index and search vast quantities of text [33]. Text mining can help collect adequate data in a timely and cost-efficiently manner [34]. Not for long, text mining based on review data collected through surveys has been used for indoor health studies particularly in workplaces (e.g., [35]). The availability of digital technologies along with easy accesses to the internet connection during the past two decades, text-mining on online reviews has become popular to access a broader range of real-world data and obtain deeper insights in various research fields [36–38].

However the use of social media as a source for big data is becoming more common [39], and its implication in design decision-making and planning is evident [40], there still exist gaps in application of such capacities in exploring occupant feedback on the IAQ during the Covid pandemic. Therefore, this research sought to investigate occupants' sentiment on Indoor Air Quality (IAQ) experience in buildings in the early Covid stay-at-home through developing NLP models based on SA computational approaches to portray perception of the building occupants against the IAQ during the time where they spent much of their time, significantly more than normal conditions. We used geo-tagged data from Twitter during the first year of the pandemic (2020) and compare it with the baseline (2019).

2.0 MATERIALS AND METHODS

We collected data from a prominent social media platform; Twitter, which freely provides historical data for academic research [41]. We used RStudio Software for the computation, which provides a user-friendly interface for programming with the R language. We developed a workflow, which entails six major steps: 1) Data Acquisition, 2) Data Cleaning 3) Text Tokenization, 4) Analysis, 5) Accuracy Evaluation, and 6) Visualization and Mapping:

2.1. Data Acquisition

In this study, we picked Twitter as the source for social media data because 1) Twitter users send short messages that are limited to 280 characters, therefore people sentiment is conveyed through a short text communication; 2) Twitter offers a free product key for academic users to access the entire historical data [23]; and 3) many studies have used Twitter as the primary social media platform for the human-centered explorations in the built environment [43]. In this research, extracted data for two consecutive years: 2019 and 2020 to compare pre and post Covid sentiments. We used the *academictwitteR* package [1], as it allows for using the API v2 key allowing for extracting historical data with minimum query restrictions from Twitter database. For performing data acquisition, we highlighted the following sets of terms with various combinations to explore users' perception about the building indoor air quality, including "air cleaner", "air filter", "air circulation", "ceiling fan", "co2 level", "control air quality"," dehumidifier", "improve air quality", "indoor air pollution", "stale air", "unhealthy air", "virus airborne", "iaq", as well as "indoor air quality" itself. These terms were queried based on exact phrases, which returned tweets that included only these terms within the posted comments to highlight IAQ scopes. To avoid bots, we only implemented acquisition based only on verified Twitter accounts.

2.2. Data Cleaning

We cleaned and structured the collected metadata, made it suitable for developing NLP models. In doing so, we removed undesired characters (e.g., URLs, sings, numbers, punctuations, emojis, etc.) and English stop words (e.g., the, is, are, he, she, etc.) from the text that cannot convey human feelings, using *dplyr* [44], *tidyr* [45], *tm* [46] and *Corpus* [47] packages together.

2.3. Text Tokenization

Tokenization is a technique applied to transform textual data into a numeric format that can then be used to train an NLP model. We performed state-of-the-art text-processing approaches by which each term mathematically is represented in binary (0,1) values leading to the matrix called Document-term Matrix (DTM) [48]. DTM allows for capturing the counts of each term within the text, allowing for applications of functions such as word frequency, word cloud, and SA in order to obtain insights from the collected data.

2.4. Analysis

In this research, we performed three types of analyses in developing NLP models, including word frequency, word cloud, and SA. Word frequency is a method to sort and visualize the magnitude of repeated words within the text –based on the created DTM in the previous step (Text Tokenization)– in a way that the frequency of words determines the order of repeated terms. Similarly, word cloud plots the words by its frequency by which the size of each plotted word represents the magnitude of the word frequency. We used *dplyr* [44] and *Wordcloud* [51] packages to implement word frequency and word cloud analyses, respectively.

Using the DTM, we developed SAs to compute occupants feedback. SA is tribute of the NLP that performs deductions on how language relates to a decision variable, extracting subjective information from narrative content. The result is either a continuous sentiment score or a positive or negative classification. These methods can mainly be divided into two groups: those that use machine learning and methods that use pre-defined dictionaries [54]. They obtain subjective insight from the occurrences of words with predetermined polarities chosen ex-ante based on experts' intuition, resulting in a technique that yields outcomes understandable to the computer. SA can be applied across levels: from a single sentence to the collection of sentences [55]. In this research, we applied pre-defined dictionaries for computing SAs using the *SentimentAnalysis* [53] package.

The sentiment lexicons propagate a numerical score, the sign of which indicates the polarity of the sentiment and the absolute value of which indicates the magnitude of the polarity. There are four popular standard sentiment lexicons including, *syuzhet*; *afinn*; *bing*; and *nrc* [56] that the target text is screened based upon one of these lexicons, considering the area or purpose of the study. The *syuzhet* lexicon gives the option to select from one of four sentiment dictionaries or create a custom dictionary. It includes 10748 words with an associated sentiment value, ranged between [−1, 1]. Negative words dominate the word pool in this lexicon since they account for 7161 out of 10748 words; positive words are the remaining 3587. The *afinn* dictionary is comprised of the internet slang and obscene words [57]. It contains 2477 words, which includes 1598 negative and 878 positive words, and a single neutral word, ”some kind”. The score is ranged between [−5, 5], much wider than that of the *syuzhet* dictionary. The *bing* dictionary [58] contains 6789 words, of which, 2006 are positive, and 4783 are negative. Table 1 lists bi-polar lexicons with corresponded number of words.

Table 1. List of bi-polar lexicons with corresponded positive and negative words and resolution values

Dictionary	Number of words	Number of positive words	Number of negative words	Resolution
syuzhet	10748	3587	7161	16
afinn	2477	879	1598	11
bing	6789	2006	4783	2

The *nrc* lexicon [58] assigns additional sentiment groups into the bi-polar (negative and positive) sentiment categories. These extra eight categories are *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, and *trust*. The *nrc* lexicon turns the score for each sentiment category from sentences rather than only the positivity and negativity scores. Table 2 lists tributes of the *nrc* lexicon with the number of sensitive words per sentiment and positive/negative polarities. Resolution is also another metric that assigns words on a continuous scale, with more extreme words having higher values. In this type of scoring, all the words in a particular letter are summed up to calculate an overall sentiment score for given location and/or time horizon. We used *nrc* lexicon for this research to compute the sentiments.

Lexicons primarily use sentiment dictionaries that include cluster of words that convey human feelings. The most frequently used dictionaries that are used in sentiment analysis include 1) the Psychological Harvard-IV (GI) dictionary 2) Henry's finance-specific (HE) dictionary [59]; 3) The Loughran-McDonald finance-specific (LM) dictionary [60]; and 4) The Quantitative Discourse Analysis Dictionary (QDAP) (Rinker, 2020). GI uses various psychological categories beyond positive and negative feelings over statements and can be retrieved through the *loadDictionaryGI()* function embedded in the *SentimentAnalysis* package. According to the psychological domain, it contains a list of 1316 positive and 1746 negative words. HE dictionary can be retrieved through the *loadDictionaryHE()* function using the same package, which is considered as a small dictionary that only includes 53 positives and 44 negative words. The LM dictionary can be retrieved from the *loadDictionaryLM()* function using the same package, which contains 145 positive and 885 negative words. The QDAP dictionary has 1280 positive and 2952 negative words, which can be called applying the *loadDictionaryQDAP()* function. In this research, we used the QDAP dictionary, as it is developed for all purposes, not targeted for a specific research domain [62].

Table 2. List of nrc sentiment categories with the corresponded number of sensitive words per category.

nrc sentiment categories with number of words per category										
Categories	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Positive	Negative
# of words	1247	839	1058	1476	689	1191	534	1231	2312	3324

2.5. Accuracy Evaluation

In this research, we evaluated the model errors based on the R-Squared, RMSE, MAE metrics.

2.6. Visualization and Mapping

In this research we performed various data visualization based on required analyses. We applied *leaflet* package for mapping geo-tagged tweets. We also used the *RColorBrewer* [52] and *ggeasy* [50] collectively package to visualize the results of word cloud analysis. We visualized results of the spatial SAs and word frequency using bar plots from the *ggplot2* [49] package. And the scatter plot from the plotly package was used to depict the temporal SAs. All these R packages are free-accessed and can be downloaded through the Comprehensive R Archive Network (CRAN) database (hyperlinked).

3.0 RESULTS AND DISCUSSION

Figure 0 maps tweets related to IAQ-related keywords posted on Twitter pages within the U.S. boundaries during 2019 and 2020. As the maps illustrate, most of tweet instances occurred in large cities and their metro areas. Similarities on the spatial pattern between the two, indicating that the users' feedback are collected from almost the same locations. Figure 1 illustrates keyword frequencies upon indoor air quality cluster keywords explored for 2020 and 2019. Contrary to the baseline (2019) that people showed their highest reactions to the indoor air quality-related subjects in July probably due to the hot weather conditions, in 2020, more reactions tweeted in September. This can be because of covid-related satisfaction/dissatisfaction tweeted in September, rather than the conventional issues such as temperature spikes. Figure 2 depicts WordCloud plots on indoor air quality-related words for 2019 and 2020, illustrating that among the captured historical posts on Twitter pages among the Twitter users in the U.S., air conditioner was the most frequent word in 2019 followed by air, hvac, fan, ceiling, heat, and ventilation on occupant feedback against indoor air quality. Indoor air quality wordcloud in 2020 shows that among the tweets, air was mentioned the most followed by ventilation, hvac, fan, ceiling, air quality, air borne, virus, mask, and air conditioner. The "ventilation", "fan", "air quality", "airborne", "virus", and "mask" have mentioned more in 2020 compared with the baseline. This probably has to do with the emergence of the covid-19 and the stay-at-home living life-style led to raises over the indoor air quality dissatisfaction.

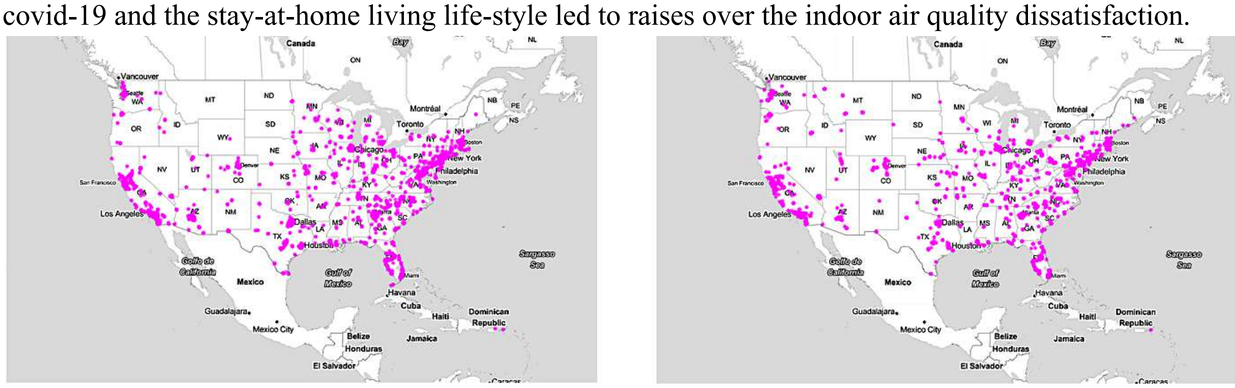


Figure 0. IAQ-related tweets and their distribution in the U.S. mainland in 2019 (left) and 2020 (right).

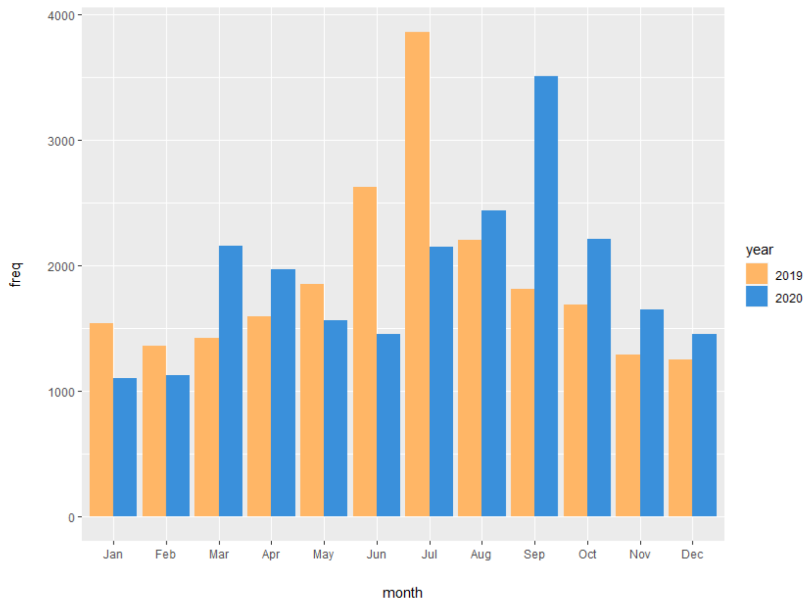


Figure 1. Keyword frequencies upon cluster keywords explored on indoor air quality for 2020 vs. 2019.

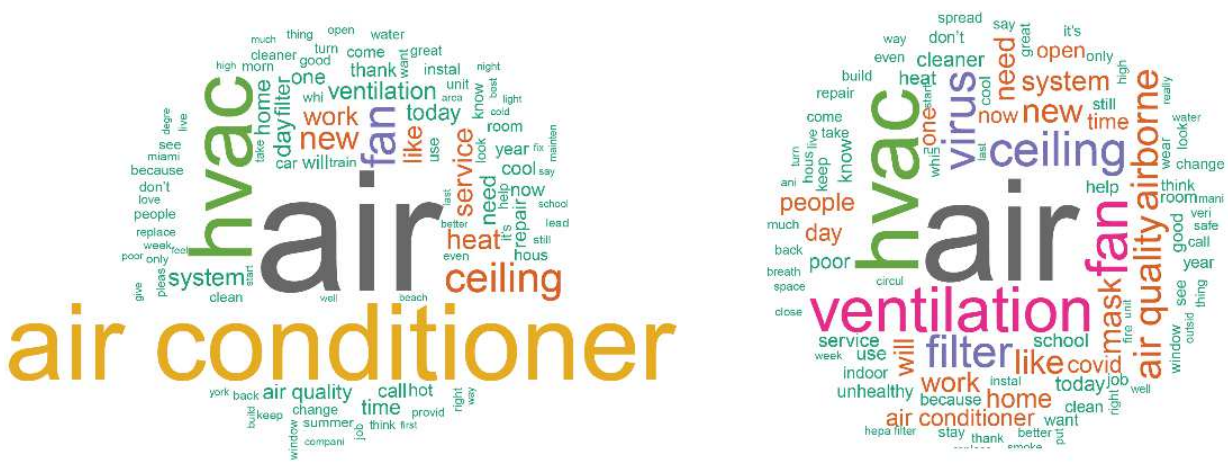


Figure 2. Wordcloud plots on IAQ-related keywords for 2019 (left) and 2020 (right).

Tables 3 and Table 4 list results of performance analysis for spatial and temporal analysis, respectively, based on the R-squared, RMSE, and MAE metrics on SAs with the combined, negativity, and positivity sentiments separately using on the *QDAP* dictionary and *nrc* lexicon for 2010 and 2019 separately. Results indicate a statistically significant performance in modeling SAs for the given information, meaning that the selected data, timeline, and geographical location on the highlighted keywords can be reliable for obtaining the feedback.

Table 3: Results of accuracy evaluation for the spatial SA using *nrc* lexicon and QDAP dictionary for 2020 and 2019 based on R-squared, RMSE, and MAE metrics.

Metric	**Sentiment**	Negativity	Positivity
2020			
R^2	**0.6465**	0.3842	0.3318
RMSE	**0.7651**	0.8938	0.7798
MAE	**0.6694**	0.793	0.6902
2019			
R^2	**0.6714**	0.4309	0.3421
RMSE	**0.7793**	0.9152	0.8029
MAE	**0.6916**	0.8243	0.7216

Figure 3 shows results of annual sentiment analyses on keywords related to the indoor air quality performed for a nationwide study in the U.S. in 2021 vs. 2020. Upon implementing the data query, 22,505 and 22,798 tweets captured for 2019 and 2020, respectively. The graph illustrates that negative sentiment scores experienced spikes in 2020 compared with the baseline (2019), meaning that the indoor air quality conditions during the lockdown probably were not satisfactory.

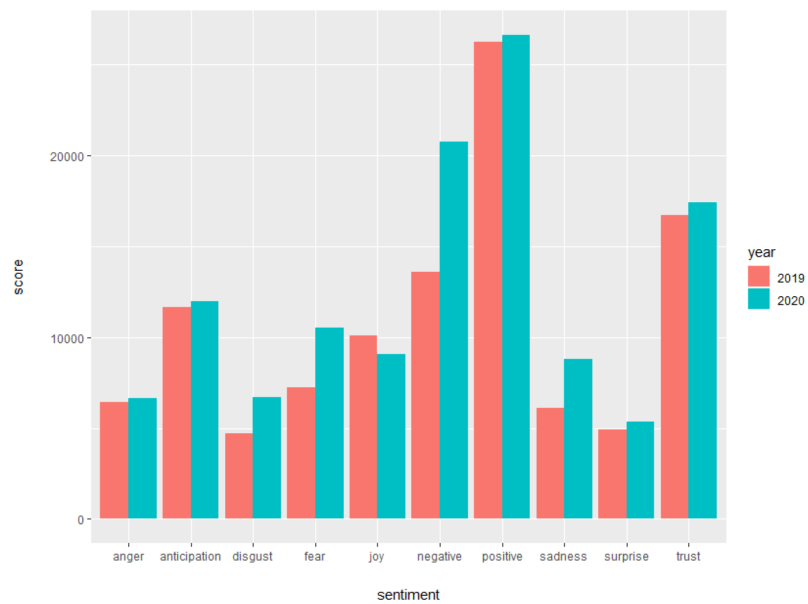
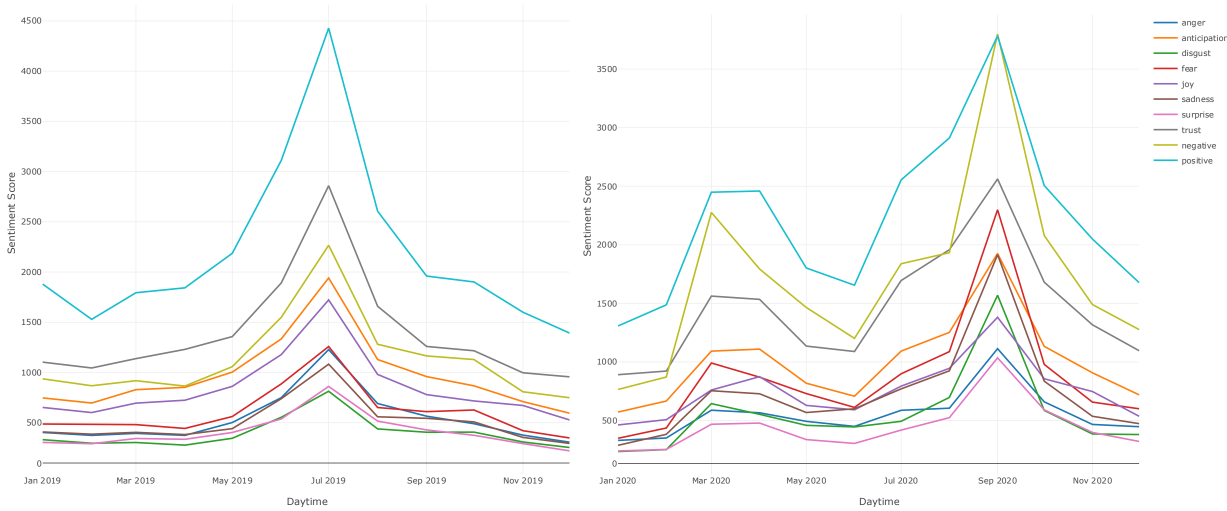


Figure 3. Spatial sentiment analysis on indoor air quality in the USA in 2019 vs. 2020.

Figure 4 shows results on temporal sentiment analysis with monthly intervals based on the QDAP dictionary and *nrc* lexicon over IAQ-related keywords in the U.S, for both 2020 and 2019. As explained earlier on Figure 1 graph, which illustrated tweet frequencies per month for the two consecutive years, the results of the monthly sentiment analysis reveal a spike on the negative sentiment in September 2020. In addition, the negative sentiment was found to be the highest in 2020, while ranked the third during the baseline year, indicating the issues people experienced during the lockdown.



(a) (b)

Figure 4. Monthly sentiment analyses on IAQ-related keywords in the U.S. during 2019 (a) and 2020 (b).

Table 4. Results of performance metrics for temporal sentiment analysis based on QDAP dictionary and *nrc* lexicon.

	Accuracy Metrics								
	R^2			**RMSE**			**MAE**		
Month	**Sentiment**	Negativity	Positivity	**Sentiment**	Negativity	Positivity	**Sentiment**	Negativity	Positivity
2020									
January	**0.6450**	0.3733	0.3149	**0.7461**	0.8815	0.7605	**0.6434**	0.7744	0.6662
February	**0.6057**	0.4096	0.2824	**0.7818**	0.9203	0.7992	**0.7017**	0.8374	0.7241
March	**0.6916**	0.4648	0.3727	**0.7738**	0.9317	0.8089	**0.6891**	0.8436	0.7314
April	**0.6579**	0.4140	0.3525	**0.7822**	0.9260	0.7984	**0.7078**	0.8507	0.7311
May	**0.6470**	0.4415	0.3218	**0.7735**	0.9167	0.7944	**0.6896**	0.8327	0.7188
June	**0.6567**	0.4195	0.3214	**0.7700**	0.9161	0.7938	**0.6827**	0.8247	0.7126
July	**0.6552**	0.4198	0.3390	**0.7796**	0.9145	0.7967	**0.6961**	0.8289	0.7203
August	**0.6514**	0.4206	0.3602	**0.7721**	0.9184	0.7945	**0.6873**	0.8307	0.7167
September	**0.7128**	0.4476	0.3843	**0.7721**	0.9278	0.8128	**0.6810**	0.8312	0.7274
October	**0.6764**	0.4461	0.3333	**0.7719**	0.9169	0.7997	**0.6827**	0.8236	0.7166
November	**0.6465**	0.3974	0.3242	**0.7681**	0.9066	0.7895	**0.6763**	0.8099	0.7030
December	**0.6692**	0.4344	0.3367	**0.7666**	0.9109	0.7886	**0.6780**	0.8184	0.7063
2019									
January	**0.6110**	0.3957	0.2854	**0.7620**	0.8997	0.7797	**0.6680**	0.8013	0.6918
February	**0.6799**	0.4035	0.3593	**0.7582**	0.9022	0.7809	**0.6633**	0.8007	0.6912
March	**0.6208**	0.3797	0.3021	**0.7663**	0.8968	0.7785	**0.6742**	0.8007	0.6930
April	**0.5997**	0.3706	0.3109	**0.7679**	0.8994	0.7778	**0.6790**	0.8065	0.6953
May	**0.6310**	0.3798	0.3140	**0.7564**	0.8910	0.7711	**0.6591**	0.7884	0.6804
June	**0.6503**	0.4146	0.3354	**0.7628**	0.9028	0.7777	**0.6736**	0.8090	0.6949
July	**0.6636**	0.3981	0.3686	**0.7522**	0.8949	0.7694	**0.6564**	0.7928	0.6797
August	**0.6521**	0.3798	0.3591	**0.7536**	0.8956	0.7684	**0.6602**	0.7965	0.6810
September	**0.6541**	0.4291	0.3315	**0.7553**	0.9027	0.7756	**0.6625**	0.8055	0.6899
October	**0.6518**	0.4102	0.3437	**0.7561**	0.9007	0.7766	**0.6612**	0.8008	0.6885
November	**0.6466**	0.3528	0.3247	**0.7609**	0.8944	0.7729	**0.6680**	0.7953	0.6851
December	**0.6636**	0.3512	0.3271	**0.7419**	0.8773	0.7580	**0.6357**	0.7624	0.6569

Figure 5(a) shows the average monthly temperature in the USA in 2019 and 2020. As the graph illustrates, the temperature spike has in July for both years, which can conform why tweets included indoor air quality-related keywords had the highest frequency (Figure 1) and accordingly the highest negative score in July in 2019 (figure 4(b)). This can also be confirmed by the average monthly energy consumption in the U.S. Figure 5-b illustrated based on data captured from the U.S. Energy Information Administration's (EIA) [63] shows that the maximum energy consumption occurred in July 2019. This can explain correlations between the maximum negative scores conveyed through the occupants' sentiments and maximum energy consumption during the same month (in 2019). According to the EIA, in 2020, 34 million people, nearly 27% of all the U.S. residents reported involving in difficulties paying their energy bills (Figure 6). This has led residents to keep their homes in an uncomfortable thermal condition, as they were not able to pay the energy bills, easily. This can contribute to the negative feedback on the indoor air quality conditions in the U.S. buildings in 2020, captured through the occupants' sentiment tweeted during the breakdown.

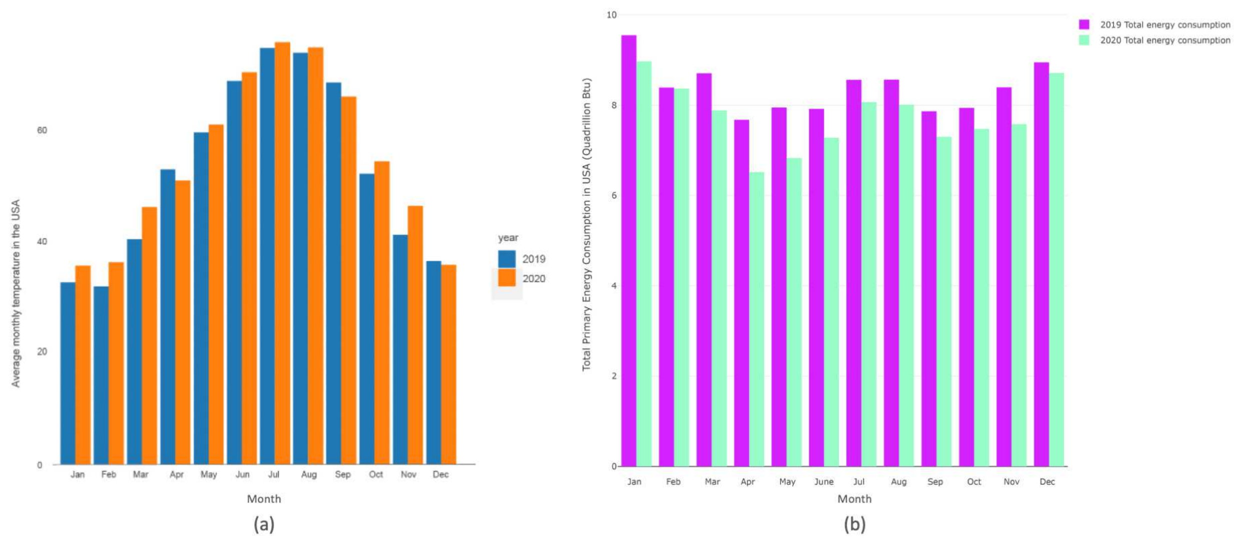


Figure 5. Averaged monthly temperature in the U.S. in 2020 and 2019 (a) and averaged monthly energy consumption in the U.S. in 2019 and 2010, Source: (EIA 2020) (b).

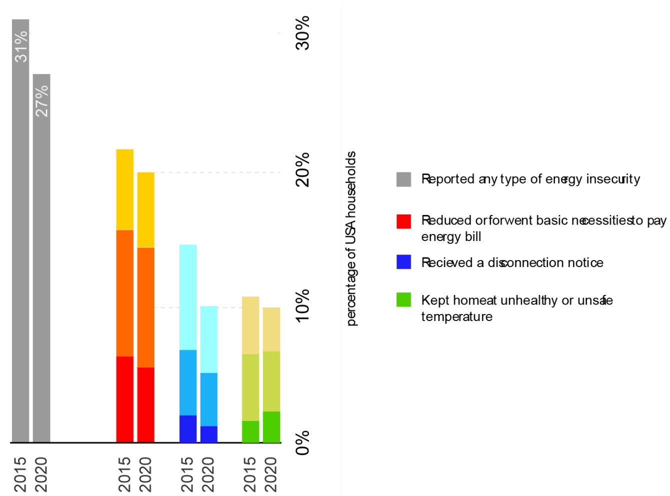


Figure 6. Percentage of the energy-burdened households based on 4 categories in the U.S. for years 2015 and 2020, Source: (EIA, 2020).

4.0. CONCLUSION

This research explored occupants' feedback on indoor air quality during the early Covid-19 stay-at-home (2020) and compared it with the baseline (2019) in order to unfold indoor environmental problems reported due to the pandemic. We extracted data from social media (herein Twitter) and applied NLP computations based on a sentiment analysis approach along with a series of data preparation processes (data cleaning and text processing) and performance simulations to implement the analytical simulations. Results suggest that the frequency of tweets upon unhealthy indoor air conditions and the negative sentiment scores both raised in the stay-at-home moments compared with the baseline. This means critical issues exist in current residential buildings across the country regarding the building's indoor environmental quality (IEQ) aspects. This research can aid architects, engineers, planners, and policymakers to better understand the impacts of lockdown on occupants' health experiences reflected directly on social media pages in order to advance design and engineering solutions on the buildings for similar scenarios, ensuring public health and resilience for future buildings, cities, and communities.

ACKNOWLEDGEMENTS

REFERENCES

[1] M.Á. Navas-Martín, I. Oteiza, T. Cuerdo-Vilches, Dwelling in times of COVID-19: An analysis on habitability and environmental factors of Spanish housing, Journal of Building Engineering. 60 (2022) 105012. https://doi.org/10.1016/j.jobe.2022.105012.

[2] American Time Use Survey - 2021 Results, (n.d.) 26.

[3] K. Schaul, B.R. Mayes, B. Berkowitz, Where Americans are still staying at home the most, Washington Post. (n.d.). https://www.washingtonpost.com/graphics/2020/national/map-us-still-staying-home-coronavirus/ (accessed September 17, 2022).

[4] C. Kakoulli, A. Kyriacou, M.P. Michaelides, A Review of Field Measurement Studies on Thermal Comfort, Indoor Air Quality and Virus Risk, Atmosphere. 13 (2022) 191. https://doi.org/10.3390/atmos13020191.

[5] World Health Organization, ed., Indoor air pollutants: exposure and health effects: report on a WHO meeting, Nördlingen, 8-11 June 1982, World Health Organization, Regional Office for Europe, Copenhagen, 1983.

[6] R.D. Brook, S. Rajagopalan, C.A. Pope, J.R. Brook, A. Bhatnagar, A.V. Diez-Roux, F. Holguin, Y. Hong, R.V. Luepker, M.A. Mittleman, A. Peters, D. Siscovick, S.C. Smith, L. Whitsel, J.D. Kaufman, Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific Statement From the American Heart Association, Circulation. 121 (2010) 2331–2378. https://doi.org/10.1161/CIR.0b013e3181dbece1.

[7] C.A. Pope, R.T. Burnett, G.D. Thurston, M.J. Thun, E.E. Calle, D. Krewski, J.J. Godleski, Cardiovascular Mortality and Long-Term Exposure to Particulate Air Pollution: Epidemiological Evidence of General Pathophysiological Pathways of Disease, Circulation. 109 (2004) 71–77. https://doi.org/10.1161/01.CIR.0000108927.80044.7F.

[8] C.A. Pope, N. Coleman, Z.A. Pond, R.T. Burnett, Corrigendum to "Fine particulate air pollution and human mortality: 25+ years of cohort studies" [Environ. Res. 183 (2020) 108924], Environmental Research. 191 (2020) 109974. https://doi.org/10.1016/j.envres.2020.109974.

[9] O. US EPA, Improving Indoor Air Quality, (2014). https://www.epa.gov/indoor-air-quality-iaq/improving-indoor-air-quality (accessed September 14, 2022).

[10] J.G. Allen, P. MacNaughton, U. Satish, S. Santanam, J. Vallarino, J.D. Spengler, Associations of Cognitive Function Scores with Carbon Dioxide, Ventilation, and Volatile Organic Compound

Exposures in Office Workers: A Controlled Exposure Study of Green and Conventional Office Environments, Environmental Health Perspectives. 124 (2016) 805–812. https://doi.org/10.1289/ehp.1510037.

[11] M. Awada, B. Becerik-Gerber, S. Hoque, Z. O'Neill, G. Pedrielli, J. Wen, T. Wu, Ten questions concerning occupant health in buildings during normal operations and extreme events including the COVID-19 pandemic, Building and Environment. 188 (2021) 107480. https://doi.org/10.1016/j.buildenv.2020.107480.

[12] T. Bidila, R.N. Pietraru, A.D. Ionita, A. Olteanu, Monitor Indoor Air Quality to Assess the Risk of COVID-19 Transmission, in: 2021 23rd International Conference on Control Systems and Computer Science (CSCS), IEEE, Bucharest, Romania, 2021: pp. 356–361. https://doi.org/10.1109/CSCS52396.2021.00065.

[13] J. Li, Y. Men, X. Liu, Z. Luo, Y. Li, H. Shen, Y. Chen, H. Cheng, G. Shen, S. Tao, Field-based evidence of changes in household $PM_{2.5}$ and exposure during the 2020 national quarantine in China, Environ. Res. Lett. 16 (2021) 094020. https://doi.org/10.1088/1748-9326/ac1014.

[14] X. Wu, R.C. Nethery, M.B. Sabath, D. Braun, F. Dominici, Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis, Sci. Adv. 6 (2020) eabd4049. https://doi.org/10.1126/sciadv.abd4049.

[15] T. Benmarhnia, Linkages Between Air Pollution and the Health Burden From COVID-19: Methodological Challenges and Opportunities, American Journal of Epidemiology. 189 (2020) 1238–1243. https://doi.org/10.1093/aje/kwaa148.

[16] an_analysis_about_transmission_modes_covid.pdf, (n.d.).

[17] A.M. Ibrahim, M.A. Hassanain, Assessment of COVID-19 precautionary measures in sports facilities: A case study on a health club in Saudi Arabia, Journal of Building Engineering. 46 (2022) 103662. https://doi.org/10.1016/j.jobe.2021.103662.

[18] H. Frumkin, COVID-19, the Built Environment, and Health, Environ Health Perspect. 129 (2021) 075001. https://doi.org/10.1289/EHP8888.

[19] T. Roh, A. Moreno-Rangel, J. Baek, A. Obeng, N.T. Hasan, G. Carrillo, Indoor Air Quality and Health Outcomes in Employees Working from Home during the COVID-19 Pandemic: A Pilot Study, Atmosphere. 12 (2021) 1665. https://doi.org/10.3390/atmos12121665.

[20] F.J. Kelly, J.C. Fussell, Improving indoor air quality, health and performance within environments where people live, travel, learn and work, Atmospheric Environment. 200 (2019) 90–109. https://doi.org/10.1016/j.atmosenv.2018.11.058.

[21] J.C. Nwanaji-Enwerem, J.G. Allen, P.I. Beamer, Another invisible enemy indoors: COVID-19, human health, the home, and United States indoor air policy, J Expo Sci Environ Epidemiol. 30 (2020) 773–775. https://doi.org/10.1038/s41370-020-0247-x.

[22] D. Sheikh Khan, J. Kolarik, P. Weitzmann, Application of an occupant voting system for continuous occupant feedback on thermal and indoor air quality – Case studies in office spaces, Energy and Buildings. 251 (2021) 111363. https://doi.org/10.1016/j.enbuild.2021.111363.

[23] R. Xie, Y. Xu, J. Yang, S. Zhang, Indoor air quality investigation of a badminton hall in humid season through objective and subjective approaches, Science of The Total Environment. 771 (2021) 145390. https://doi.org/10.1016/j.scitotenv.2021.145390.

[24] X. Zhou, Y. Liu, C. Song, J. Liu, Impact of Temperature and Microstructure on the Emission Characteristics of VOC in Porous Building Materials, Procedia Engineering. 121 (2015) 1067–1075. https://doi.org/10.1016/j.proeng.2015.09.104.

[25] H. Jarnstrom, K. Saarela, P. Kalliokoski, A. Pasanen, Reference values for indoor air pollutant concentrations in new, residential buildings in Finland, Atmospheric Environment. 40 (2006) 7178–7191. https://doi.org/10.1016/j.atmosenv.2006.06.021.

[26] H. Tang, Y. Ding, B.C. Singer, Post-occupancy evaluation of indoor environmental quality in ten nonresidential buildings in Chongqing, China, Journal of Building Engineering. 32 (2020) 101649. https://doi.org/10.1016/j.jobe.2020.101649.
[27] X. Guo, K. Lee, Z. Wang, S. Liu, Occupants' satisfaction with LEED- and non-LEED-certified apartments using social media data, Building and Environment. 206 (2021) 108288. https://doi.org/10.1016/j.buildenv.2021.108288.
[28] G. Chinazzo, Geographical distribution of the sources of IEQ complaints: an analysis through text-mining of online job reviews, J. Phys.: Conf. Ser. 2069 (2021) 012176. https://doi.org/10.1088/1742-6596/2069/1/012176.
[29] H. Villeneuve, W. O'Brien, Listen to the guests: Text-mining Airbnb reviews to explore indoor environmental quality, Building and Environment. 169 (2020) 106555. https://doi.org/10.1016/j.buildenv.2019.106555.
[30] J. Silge, D. Robinson, Text Mining with R, n.d. https://www.tidytextmining.com/.
[31] K.B. Cohen, ing and Text Mining Biomedical Natural Language Process-, (n.d.) 24.
[32] S. Green, J. Heer, C.D. Manning, Natural Language Translation at the Intersection of AI and HCI, ARTIFICIAL INTELLIGENCE. (n.d.) 13.
[33] P.M. Nadkarni, L. Ohno-Machado, W.W. Chapman, Natural language processing: an introduction, J Am Med Inform Assoc. 18 (2011) 544–551. https://doi.org/10.1136/amiajnl-2011-000464.
[34] H.S. Abdulaali, I.M.S. Usman, M.M. Hanafiah, Online review analysis of perceived Indoor Environmental Quality (IEQ) at former green certified hotels in Kuala Lumpur, Malaysia, (2020) 14.
[35] M. Moezzi, J. Goins, Text mining for occupant perspectives on the physical workplace, Building Research & Information. 39 (2011) 169–182. https://doi.org/10.1080/09613218.2011.556008.
[36] R. Romero-Silva, S. de Leeuw, Learning from the past to shape the future: A comprehensive text mining analysis of OR/MS reviews, Omega. 100 (2021) 102388. https://doi.org/10.1016/j.omega.2020.102388.
[37] D. Antons, E. Grünwald, P. Cichy, T.O. Salge, The application of text mining methods in innovation research: current state, evolution patterns, and development priorities, R&D Management. 50 (2020) 329–351. https://doi.org/10.1111/radm.12408.
[38] University of Liechtenstein, S. Debortoli, O. Müller, IT University of Copenhagen, I. Junglas, Florida State University, J. vom Brocke, University of Liechtenstein, Text Mining for Information Systems Researchers: An Annotated Topic Modeling Tutorial, CAIS. 39 (2016) 110–135. https://doi.org/10.17705/1CAIS.03907.
[39] S. Bazzaz Abkenar, M. Haghi Kashani, E. Mahdipour, S.M. Jameii, Big data analytics meets social media: A systematic review of techniques, open issues, and future directions, Telematics and Informatics. 57 (2021) 101517. https://doi.org/10.1016/j.tele.2020.101517.
[40] P. Furnham, Automation and Autonomy of Big Data-driven Algorithmic Decision-M aking, (n.d.) 7.
[41] A.E. Kim, H.M. Hansen, J. Murphy, A.K. Richards, J. Duke, J.A. Allen, Methodological Considerations in Analyzing Twitter Data, JNCI Monographs. 2013 (2013) 140–146. https://doi.org/10.1093/jncimonographs/lgt026.
[42] Getting Started with the Twitter API, (n.d.). https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api (accessed November 30, 2022).
[43] L. Pilař, L. Kvasničková Stanislavská, J. Pitrová, I. Krejčí, I. Tichá, M. Chalupová, Twitter Analysis of Global Communication in the Field of Sustainability, Sustainability. 11 (2019) 6958. https://doi.org/10.3390/su11246958.
[44] H. Wickham, R. François, L. Henry, K. Müller, dplyr: A Grammar of Data Manipulation, 2022. https://dplyr.tidyverse.org.
[45] H. Wickham, M. Girlich, tidyr: Tidy Messy Data, 2022. https://tidyr.tidyverse.org, https://github.com/tidyverse/tidyr.
[46] I. Feinerer, K. Hornik, tm: Text Mining Package, 2020. https://CRAN.R-project.org/package=tm.

[47] L. Huang, P. O. Perry, F. Årup Nielsen, M. Porter, R. Boulton, corpus: Text Corpus Analysis, 2021. https://leslie-huang.github.io/r-corpus/.

[48] Y. Xie, L. Le, Y. Zhou, V.V. Raghavan, Deep Learning for Natural Language Processing, in: Handbook of Statistics, Elsevier, 2018: pp. 317–328. https://doi.org/10.1016/bs.host.2018.05.001.

[49] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York, 2016. https://ggplot2.tidyverse.org.

[50] J. Carroll, A. Schep, J. Sidi, B. Rudis, M. El Fodil Ihaddaden, T. Neitmann, Easy Access to “ggplot2” Commands, 2021. https://github.com/jonocarroll/ggeasy.

[51] I. Fellows, wordcloud: Word Clouds, 2018. http://blog.fellstat.com/?cat=11.

[52] E. Neuwirth, RColorBrewer: ColorBrewer Palettes, 2022. http://colorbrewer2.org/.

[53] N. Proellochs, S. Feuerriegel, SentimentAnalysis: Dictionary-Based Sentiment Analysis, 2021. https://github.com/sfeuerriegel/SentimentAnalysis.

[54] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams Engineering Journal. 5 (2014) 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011.

[55] N. Leelawat, S. Jariyapongpaiboon, A. Promjun, S. Boonyarak, K. Saengtabtim, A. Laosunthara, A.K. Yudha, J. Tang, Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning, Heliyon. 8 (2022) e10894. https://doi.org/10.1016/j.heliyon.2022.e10894.

[56] M. L. Jockers, Syuzhet: Extract Sentiment and Plot Arcs from Text, 2015. https://github.com/mjockers/syuzhet.

[57] F.Å. Nielsen, A new ANEW: Evaluation of a word list for sentiment analysis in microblogs, (2011). http://arxiv.org/abs/1103.2903 (accessed October 18, 2022).

[58] source_A-Proposed-Approach-for-Emotion-Lexicon-Enrichment.pdf, (n.d.).

[59] E. Henry, Market Reaction to Verbal Components of Earnings Press Releases: Event Study Using a Predictive Algorithm, Journal of Emerging Technologies in Accounting. 3 (2006) 1–19. https://doi.org/10.2308/jeta.2006.3.1.1.

[60] T. Loughran, B. Mcdonald, Textual Analysis in Accounting and Finance: A Survey, Journal of Accounting Research. 54 (2016) 1187–1230. https://doi.org/10.1111/1475-679X.12123.

[61] T.W. . Rinker, Quantitative Discourse Analysis Package, 2020. https://github.com/trinker/qdap.

[62] S. Taylor, V. Keselj, Using Extractive Lexicon-based Sentiment Analysis to Enhance Understanding of the Impact of Non-GAAP Measures in Financial Reporting, (n.d.) 7.

[63] Homepage - U.S. Energy Information Administration (EIA), (n.d.). https://www.eia.gov/index.php (accessed November 30, 2022).