



# Energy in the backseat? Investigating decarbonization dialogue in supply chain tweets during and after COVID-19

Umer Shahzad<sup>1</sup> · Tuhin Sengupta<sup>2</sup> · Amar Rao<sup>3</sup> · Gagan Deep Sharma<sup>4,5,6</sup>

Received: 29 November 2022 / Accepted: 19 December 2023  
© The Author(s) 2024

## Abstract

While we move into the seventh year of the signing of Paris agreement, research scholars and supply chain firms have paid a lot of emphasis on environmental sustainability with the aim of achieving net zero targets by 2050. However, the global pandemic has somewhat disturbed the focus from environment to resilience due to severe economic implications of COVID-19. In this paper, we contribute to the very scant discussion on Twitter Analytics by analysing supply chain tweets with COVID-19 at the backdrop. Our approach involves analysing how decarbonization related discussions have evolved by capturing the tweets across three timelines: pre pandemic, pandemic and post pandemic. By integrating descriptive analytics, content analytics and machine learning algorithm in topic modelling, we extract textual intelligence related to emissions and pollution from leading firms involving supply chain management. We find that although decarbonization related discussions are at bare minimum in terms of the proportion of discussions within the supply chain context, the overall emotion of tweets indicate fear across all three timelines. Moreover, it was surprising to note that although pollution levels came down due to low economic activity during pandemic, we found more discussions during COVID in comparison to pre-COVID times. Pollution and waste caused by plastics, fuel consumption, reduction in greenhouse gas emission are some of

---

✉ Gagan Deep Sharma  
angrishgagan@gmail.com

Umer Shahzad  
Shehzad-umer@hotmail.com

Tuhin Sengupta  
tuhin.sengupta@iimranchi.ac.in

Amar Rao  
amarydvrao@gmail.com

<sup>1</sup> Brandenburg University of Technology, Chair of Energy Economics, Cottbus, Germany

<sup>2</sup> Department of Operations Management, Indian Institute of Management Ranchi, 5th Floor, Suchana Bhawan, Meur's Road, Audrey House Campus, Jharkhand 834004, India

<sup>3</sup> BML Munjal University, 67th KM Stone, NH-8, Gurugram, Haryana 122413, India

<sup>4</sup> Széchenyi István University of Győr, Győr, Hungary

<sup>5</sup> University School of Management Studies, Guru Gobind Singh Indraprastha University, New Delhi, India

<sup>6</sup> Lebanese American University, Beirut, Lebanon

the key topics that emerged during pandemic times. Our paper makes a modest contribution on the role of social media analytics within supply chain context around COVID-19.

**Keywords** Supply chain · Pollution · Fuel consumption · Twitter · Machine learning · Sustainability

## 1 Introduction

Decarbonization is a term which commonly refers to the reduction of carbon dioxide output to the atmosphere (Deloitte, 2021). In other words, decarbonization is a process through which carbon emissions are reduced by deploying low carbon sources in the entire supply chain. According to a recent report from KPMG, organizations (mainly Fortune 500 companies) have already outlined decarbonization strategies at the backdrop of net-zero imperative (KPMG, 2021). The net-zero imperative is motivated by the 2015 Paris Agreement with the intention to pursue carbon neutrality by 2050 (Meinshausen et al., 2022; Wu et al., 2022). One critical strategy should focus on the strategic foresight in terms of resolving competing imperatives towards decarbonization. For instance, while business leaders are promoting net-zero activities, firms are still struggling to recover from post pandemic phase, create a long-term vision for the rapidly growing technological space which includes automation, disruptive technology including AI, in addition to increasing competitive markets much attributed to change in focus towards consumer centric responsive supply chains (Example: Amazon Online Marketplace) which are more digitally aware in order to serve customers. This points us to the importance of aligning business strategy towards clean energy sources, rewarding senior executives for environmental performance and reporting tangible progress in emission reduction initiatives. These initiatives will apply across different stages in the supply chain which includes sourcing strategies, manufacturing processes, distribution networks including logistics, transportation and warehousing. Hence, a process driven mechanism to reliably report such initiatives not only improve overall credibility but also increases regulatory agility in western markets where regulations involving emissions are strict. For instance, the Climate Mobilizing Act enacted by New York City in 2019 to cap carbon emissions may result in billions of dollars in fines for the real estate sector if they fail to decarbonize (KPMG, 2021).

A recent paper finds four common barriers to supply chain decarbonization: lack of awareness, resistant mindset, major upfront investment cost, and lack of expertise (Zhang et al., 2022). Hence, the next set of strategies to counter these mentioned barriers must include climate focused partnerships. Partnerships not only enhance innovation in supply chains (Xie et al., 2022) but also could lead to cost sharing contracts in different greening initiatives (Dai et al., 2017; Wang et al., 2020). It has been documented that innovation does improve the environmental quality (Appiah et al., 2023). Also, cost sharing contracts would not only reduce the problem of large investments borne by a single firm but also enhance the expertise level as multiple stakeholders would come together to address the problem. For instance, Hotel Hilton in Bali, Indonesia was recognized for their effort in eliminating more than 460,000 single use water bottles and realized 40% cost savings by partnering with a local Bali based firm (KPMG, 2021). With regard to the resistant mindset and lack of awareness issue, firms need to build trust among different stakeholders and deploy technology such as machine learning (Feng et al., 2022), blockchain (Karim et al., 2023; Kouhizadeh & Sarkis, 2018) to drive actionable insights across the entire value chain. In short, establishing credibility is the

first step towards building trust followed by verifiable data instead of greenwashing (Delmas & Burbano, 2011) to impress across stakeholders.

While the argument on the importance of decarbonization is well documented in both academic and grey literature, there has been a growing concern on how firms will evolve post COVID-19 pointing us towards the conflicting imperative towards achieving net-zero level. It is well observed that as businesses have started recovering and economic activities have increased, unsustainable consumption practices have increased pollution levels during the recovery phase (Dutheil et al., 2021) which will eventually impact resources affecting a very large proportion of the global population especially in emerging markets. According to a recent report in Ernst and Young, on how European Union could revive both economically and environmentally post pandemic, a three-pronged approach, which includes financial stimuli towards resilient economy, fiscal incentives to mitigate climate change and focus on circular economy, could positively transition Europe towards their emission reduction goals (Ernst & Young, 2022). However, due to its recency effect, we have very little information both in the academic literature and grey literature on how firms are evolving in line with their net-zero goals while recovering economically post pandemic. While, we always argued in favour of triple bottom line approach, the current pandemic has shown how difficult it is to focus on all three dimensions of sustainability. While economic recovery is something firms cannot ignore, it has to be seen how environmental concerns and their strategies evolved throughout this phase. The overarching objective of this study is to understand how discussions have evolved concerning pollution in supply chains from pre-COVID to post-COVID phase since focusing on pollution is counter intuitive to the immediate objective of economic recovery during post-COVID at the firm level. While a qualitative approach would have been suitable, as firms' responses towards resilience and environment evolve, we believe that such an approach would restrict us in capturing a wide variety of discussions across industrial sectors. We, therefore, decided to collect data from Twitter due to three important reasons. First, at a generic level, social media analytics have proved to be instrumental during the global pandemic in varying applications such as government crisis management (Chon & Kim, 2022), mitigating the negative effects of COVID-19 (Zhang et al., 2020), understanding the emotions of stakeholders impacted by the pandemic (Tinguely et al., 2020). Second, Twitter has shown its ability to highlight supply chain issues (Schmidt et al., 2020). Third, Twitter allows us to collect a large data set which could capture the issues concerning pollution at the firm level from a temporal perspective. This provides us with specific advantages in understanding how priorities or strategies (in the form of discussions/tweets) have evolved from pre pandemic to post pandemic concerning pollution. We present our research gaps as follows:

**RQ 1** What are the emerging discussions on decarbonization in supply chains in pre pandemic times? How has the discussion evolved during and recovery phase of COVID-19.

Before the COVID-19 pandemic, discussions on decarbonization in supply chains were already gaining traction, as businesses and governments recognized the environmental impact of global trade and production. Consumers were becoming more conscious of the environmental impact of the products they purchased. As a result, there were discussions about the need for greater transparency and traceability in supply chains, allowing consumers to make informed choices and encouraging companies to adopt sustainable practices. Businesses were recognizing the importance of engaging their suppliers in decarbonization efforts. Collaborating with suppliers to improve their environmental performance could lead to a ripple effect throughout the supply chain, reducing emissions and promoting sustainability. The

circular economy concept, which focuses on reducing waste and maximizing resource utilization, was becoming a prominent topic in supply chain discussions. However, it needs to be seen whether Twitter data continue to provide such valuable discussion during and after the pandemic.

Our methodological approach involves various steps. First, we collect tweets based on a certain search criterion. Then, we cleaned the data and present the descriptive statistics. We apply three different techniques to answer our research questions: topic modelling, emotion analysis and sentiment analysis by naïve bayes algorithm, a supervised machine learning technique. COVID-induced sentiment has been argued to influence intraday volatility spillovers between energy and exchange traded funds (Naeem et al., 2023). Our findings reveal that decarbonization efforts in the post pandemic recovery phase was absent and there was no discussion on any linkages potential strategies towards reducing pollution in the long run. This is concerning for leading supply chain firms and environmentalists on the use and engagement of social media platforms towards highlighting the importance of decarbonization activities at various stages within the supply chain. Further, we did not find any focused involvement of stakeholder engagement initiatives towards decarbonization as pointed out by one of the papers in the literature. Our papers make a modest attempt to point out how the global pandemic may have diverted attention from decarbonization to resilience thereby risking net zero targets of nations across the globe with far reaching implications to sustainability.

The remainder of the section is as follows: Sect. 2 presents the review of related literature ranging from supply chain, pollution, twitter and COVID-19. The research gap is derived by carefully reviewing the above-mentioned literature streams. The third section presents the methodology and the data description. Here, in addition to providing the theoretical understanding of different methods that we have adopted in the paper, we also explain the data collection process from Twitter, followed by descriptive analysis of the data. Section 4 presents the results and findings of the topic modelling, sentiment analysis and emotions across different timelines from pre-covid to post covid. We conclude the paper in the final section by highlighting the core findings, implications to theory and practice followed by limitation and future scope of our work.

## 2 Literature review and theoretical framework

The review of literature is divided into four streams. First, we base our argument by discussing how the theory on social media analytics have evolved into different methods and business function applications. Second, we discuss the current studies in the interface of supply chain and pollution as this remains our core focus. Then, we further capture the discussions on supply chains in twitter since our sample involves tweets. Second, we explain the related literature on environmental pollution in the context of the pandemic since our focus also entails COVID-19. We draw our research gap by triangulating the information synthesized from these streams of literature.

### 2.1 Theoretical framework

The underlying theoretical discussion is based on the application areas of social media analytics (Rathore et al., 2017). Social media analytics is currently used in various businesses to capture insights which could aid performance and productivity issues in various domains. The theory specifically points out how user-oriented interaction processes such as discussion

and debates in social media platforms such as Twitters have evolved over the last decade (Kaplan & Haenlien, 2010). Such user-oriented content is critical as both direct and indirect stakeholders of businesses often engage in issues that warrants opinions from different users. With regard to methods, domains and objectives, Rathore et al. (2017) points out how different data mining objective such as classification, clustering, association rule discovery, regressions have been applied in different industry such as public administration, consumer discretionary, education, healthcare among others. A careful examination of the business function would suggest emergency management, operations, accounting and marketing, stakeholder engagement as the prominent areas of business function applications. As users generated content is evolving at a faster pace, text analytics, visualization and network analytics have been used extensively in such business domains. With specific reference to industry wise analysis, the authors find sentiment analysis, topic modelling as one of the most commonly used techniques to develop business insights. We base our understanding and way forward in social media analytics and apply to supply chain issues which is currently an emerging domain in developing insights into key issues. In this paper, we explore pollution as a critical focal point keeping in mind net zero targets. However, with the global pandemic disturbing the net zero focus, we wish to explore how supply chain discussions revolving around pollution has evolved across the pandemic timelines.

## 2.2 Supply chain, pollution and twitter

Studies related to supply chain and pollution are focused into three broad areas, namely, regulatory and policy measures leading to pollution reduction, role of technology and stakeholders towards pollution mitigation and impact of strategic responses and consumer pressure. Ouardighi et al. (2016), through a game theoretic approach, determine how the role of double marginalization towards the trade-off between pollution emissions and abatement activities in a vertical supply chain. On similar lines, Wang et al. (2022), using a Stackelberg model, analyse a two-period supply chain where manufacturer participates in a cap-and-trade scheme with an uncertain emission permit price, while the retailer faces a price sensitive demand. The author finds that the abatement level increases with the permit price and decreases with uncertainty. It also highlights that when the toxicity of the pollutants is low, the impact of emission trading on social welfare may be stronger thus highlighting the role of such abatement strategies towards social welfare. Sim and Kim (2021) analyse a scenario where manufacturer is polluting the environment through production activities and therefore determine abatement efforts while the government imposes emission penalties. The authors find that imposing a fraction of penalties on the retailer might reduce manufacturer emissions. Chung et al. (2013) highlights the importance of firms' decisions on supply chain networks. In response to one firms' decision in price change and to environmental taxes, the manufacturer may strategically change many decisions ranging from shipment patterns, inventory management, the decision of paying environmental taxes versus recycling decisions etc. Then, there are specific technology deployment studies concerning pollution in supply chains. For instance, Niu et al. (2021) finds that environmental sustainability within a supply chain improves when supply chain adopts blockchain technology and supply chain uncertainty is small. Kraines (2002) developed an integrated system trade-off model to assess cost and pollution associated with transportation in the coke making supply chain in China. The author finds that transfer of low pollution truck technology results in high pollution reduction thus highlighting the importance of truck transportation in coke making industry.

Studies related to twitter data and supply chain are scant and diverse in terms of its focus. Chae et al. (2015) was one of first few papers that analysed supply chains from different perspective. The authors touched upon issues ranging from information sharing, need for hiring professionals, stakeholder communication, risks, corporate social responsibility including environmental standards and disruptions in supply chains. Responses in twitter concerning supply chains are critical to a wider audience. Schmidt et al. (2020) confirmed this hypothesis by exploring the twitter response to supply chain glitches on stock market returns. This confirms the relevance of social media platforms such as Twitter and its ability to elevate the overall importance of supply chain related problems. Shifting our discussion on supply chain challenges, Sharma et al. (2020), by relying on twitter data from NASDAQ 100 firms, found that firms are facing challenges in terms of deploying technology, the issue of matching supply with demand and most importantly the strategic need to makes supply chains more resilient (Janjua et al., 2021). Interestingly, in line with our context, the authors also highlight that firms are experiencing challenges in implementing sustainability within their supply chains. A previous study seems to be complementing the issue raised on the pretext of sustainability (Mishra & Singh, 2018). Here, the authors, developed waste minimizing strategies by recommending a framework for execution process by using an example for beef supply chains. Although the tweets involved large data sets concerning beef products, we may argue that such strategies could be developed for other product types depending on the severity of its impact on environment.

### 2.3 Environmental pollution & COVID-19

The literature on environmental pollution and COVID-19 is fast evolving. However, our intention is to highlight those studies which show how pollution levels have changed during or post COVID-19. Yang et al. (2021) shows that COVID-19 improved PM 2.5, PM 10, NO<sub>2</sub>, and CO levels but did not reduce SO<sub>2</sub> and O<sub>3</sub> levels. Further, surface water, coastal water and groundwater became more cleaner as a result of COVID-19. The authors attribute travel restrictions as the main reason for carbon emission reduction. This is similar to Razzaq et al. (2020) except for the fact that in the latter, the author found that in high polluted areas, the intensity of COVID-19 was more in states within USA. In the case of Egypt, Mostafa et al. (2021) found that NO<sub>2</sub>, CO and GHG emissions decreased while ozone level increased. However, the author warns that post lifting of lockdown, such beneficial impacts may fade away unless stricter environmental laws are adopted and implemented. Other studies also show similar results in different contexts thus highlighting the impact of COVID-19 and pollution (Kephart et al., 2021; Keshtkar et al., 2022; Marwah & Agrawala, 2022). Hence, it is fairly evident that pollution levels came down due to the global pandemic. It is therefore, interesting to understand the association post covid and the underlying reasons if pollution levels remain reduced. This leads to our motivation of understanding how supply chains must change or changing towards reducing their carbon footprint. We argue our research gaps in the next section by summarizing the learnings from the literature.

## 3 Research gaps

We wish to present couple of observations from the review of related literature above. First, although studies concerning supply chain and pollution is evolving, we found very little evidence on *how supply chains are tackling pollution* apart from some game theoretic models

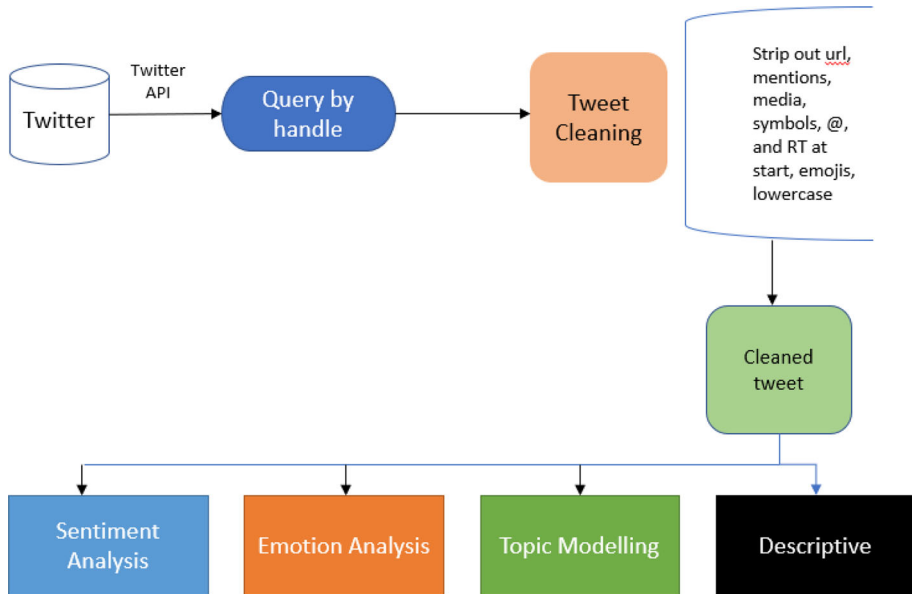
explaining a theoretical model under different regulatory conditions. Second, studies concerning pollution levels post covid is currently absent in the literature. Although, it is evident that pollution levels will rise, however how organizations, firms or supply chains would act or react in the recovery stage of pandemic specially when economic activities will pick up is currently missing in the literature. Third, although it is well established that twitter responses are critical in supply chain issues, there is very little knowhow how those responses are towards pollution abatement and mitigation strategies especially when the world is recovering from the pandemic. We do know that twitter responses have highlighted the problem of environmental sustainability in supply chains. However, the severity of the issue and broad discussion themes on environmental sustainability is currently missing when it comes to understanding how social media is responding in the supply chain context. The objective of our paper is a modest attempt to understand and compare how twitter responses have emerged from pre-covid to post-covid times in the supply chain context thus shedding light on how environmental sustainability can be achieved across different supply chain stages. The outcome of our work would also lay foundation on the discussion on how organizations should move towards stricter pollution-controlled activities within their firms while maintaining economic recovery from the global pandemic.

## **4 Data and methodology**

### **4.1 Collection of data from Twitter**

The objective of this research is to conduct an exploratory analysis on a curated set of tweets emanating from the official Twitter accounts of organizations involved in supply chain management. The analysis incorporates techniques such as frequency analysis, sentiment analysis, graph theory, and topic modeling. The overall methodology is delineated into three discrete stages: data acquisition, data preprocessing, and data analysis, as illustrated in Fig. 1.

The initial phase involves collecting tweets from organizations listed in the Forbes 2000 and Gartner's Top 50 supply chain companies. Utilizing Twitter's streaming APIs, we achieved low-latency access to a global dataset that includes tweet text along with associated metadata, such as posting time and geographical coordinates (when geolocation is enabled). Specifically, we employed Twitter's Standard Search API v1.1 to conduct simple queries against the indices of recent or popular tweets. This functionality approximates, albeit not identically, the search feature available in Twitter's mobile and web interfaces. Data collection spans three distinct timelines: pre-COVID (January 2019 to January 2020), during COVID (February 2020 to December 2021), and post-COVID (January 2022 to June 2022). The selected time periods for data collection—pre-COVID (January 2019 to January 2020), during COVID (February 2020 to December 2021), and post-COVID (January 2022 to June 2022)—are strategically chosen to offer a comprehensive understanding of the evolution of discourse on decarbonization within supply chains. The pre-COVID phase serves as a baseline, representing industry discussion and sentiment prior to the global pandemic. The during-COVID phase captures the immediate and lingering effects of the pandemic, a period when supply chain resilience became a predominant focus, potentially overshadowing sustainability goals. The post-COVID phase seeks to understand whether the discussion returns to decarbonization as firms transition to recovery and potentially prepare for a more sustainable future. These timeframes are not only relevant for capturing shifts in industry priorities



**Fig. 1** Flowchart of the dataset preparation and analysis

and public sentiment but are also critical for providing actionable insights for policymakers and supply chain firms.

Upon retrieval, tweets were filtered based on specific keywords such as “emission,” “emissions,” and “pollution.”

The rationale for this selective approach is multifold. First, these terms are directly associated with the core subject of decarbonization, making them apt indicators for gauging discourse in this context. Second, they are commonly used terminology in both layman and scientific discussions surrounding environmental sustainability, thereby maximizing the likelihood of capturing a wide range of conversations and viewpoints. Third, these terms intersect with critical aspects of supply chain operations, such as transportation and manufacturing processes, that are significant contributors to carbon emissions and pollution. By focusing on these keywords, the study aims to extract meaningful and relevant data that can illuminate how supply chain companies are publicly discussing, or perhaps notably not discussing, crucial issues related to decarbonization.

A subset of the dataset, comprising 2,215,405 tweets, is elucidated in Table 3. Prior to any analytical procedures, it is imperative to cleanse the dataset to eliminate noise and enhance interpretability. Given that tweets often contain informal language, URLs, and emojis, a “sanitized” version of the text was generated. This entailed the removal of hashtags, user mentions, URLs, media, and symbols. To maintain purity of the data, retweets indicated by “RT” and other extraneous symbols were also expunged. The remaining textual content was converted to lowercase, and tweets not in English were translated using the TextBlob Python module. In the analytical phase, sentiment analysis was executed on the cleansed dataset. For this purpose, we employed the TextBlob package, which bases its sentiment analysis on a predefined lexicon. The package provides two metrics: polarity, ranging from -1.0 to 1.0, and subjectivity, with a score between 0.0 and 1.0. Each tweet was assigned two sentiment scores: one reliant solely on the textual content and another incorporating both text and



emojis. Emojis were identified using a comprehensive Unicode list and were translated to their official or commonly recognized names for accurate sentiment interpretation.

## 4.2 Descriptive analysis

Twitter data contain a great deal of information, such as tweets and metadata (e.g., user information). The descriptive analysis focuses on descriptive data, such as the count of tweets, likes, and retweets, the distribution of tweet types, and the number of hashtags.

The distinction is in the number of metrics. While a modest number of metrics (such as sample size, response rate, and respondent profile) are utilized for survey data, the richer nature of Twitter data enables intelligence extraction by utilizing a huge number of metrics on tweets, users, hashtags, and URLs, etc.

## 4.3 Content analysis

The nature of social media data is “unstructured” because it consists mostly of text. Therefore, it is vital to utilize content analytics (CA), which refers to a wide range of natural language processing (NLP) and text mining techniques, in order to extract intelligence from Web 2.0 (Chau & Xu, 2012). The text of a tweet is made of a short list of words, hashtags, URLs, and other information. Intelligence gathering, therefore, requires careful consideration of text cleansing and processing (e.g., “Hector, a Delivery Operations Manager, is determined to drive environmental change at Amazon. Heâ€™s proud to be part of the team working to make zero-emission delivery a reality. See Hectorâ€™s story: <https://t.co/DSdhgdcXsb> <https://t.co/go4FUgxWJ8>”) contains not just text, but urls, trademark signs, hashtags, etc.). Therefore, specialized text mining techniques, such as sentiment analysis, are necessary to extract such opinions. Algorithms for text mining and machine learning are crucial components of CA. Text mining turns unstructured texts (or documents) into formatted data (or documents) using techniques such as tokenization, n-grams, stemming, and the removal of stop words (extraneous words) (c.f., Weiss et al., 2010). Using machine learning algorithms such as clustering and association analysis, these modified texts can be utilized for text summarization, keyword analysis, word frequency analysis, and text clustering. Despite the fact that the presence of CA in supply chain research (Georgi et al., 2010; Seuring & Gold, 2012; Vallet-Bellmunt et al., 2011), the approach has been manual or semi-manual, relying heavily on human interpretations. Due to the huge data nature of Twitter data, CA in TA depends on text-processing techniques and algorithms.

In CA, word analysis is the beginning point. It comprises frequency analysis of terms, document summary, and grouping. In information retrieval, term frequency (TF) is commonly employed. The combination of TF and n-gram enables the identification of significant terms within documents. Overall, our studies aid in identifying debate themes and reveal distinct Twitter usage based on emissions related tweets by supply chain’s organizations. These highlighted topics are valuable for document-level analysis (e.g., clustering) using unsupervised machine learning methods.

Term Frequency (TF): The term frequency is a measure of how frequently a term occurs in a document. It is calculated as follows:

$$TF(t, d) = (\text{Number of times term } t \text{ appears in document } d) / (\text{Total number of terms in document } d)$$

where:

t is a term

d is a document

N-grams: N-grams are sequences of n words from a text. For example, the bigram "supply chain" is an example of a 2-g. N-grams are used to identify significant terms within documents.

TF-IDF: TF-IDF (Term Frequency-Inverse Document Frequency) is a weighting scheme used to identify the importance of a term in a document. It is calculated as follows:

$$TF - IDF(t, d) = TF(t, d) * \log(N/DF(t))$$

where

N is the total number of documents in the corpus.

DF(t) is the number of documents in the corpus that contain the term t.

Cosine Similarity: Cosine similarity is a measure of similarity between two vectors, typically used to compare the similarity between two documents. It is calculated as follows:

$$\text{cosine\_similarity}(d1, d2) = (d1.d2) / (||d1|| * ||d2||)$$

where

d1 and d2 are two document vectors.

. represents the dot product of the two vectors

$$||d1|| \text{ and } ||d2|| \text{ are the lengths of the vectors}$$

Overall, the use of frequency analysis of terms, including TF, n-grams, TF-IDF, and cosine similarity, can help to identify significant terms within documents and aid in the identification of themes and patterns within social media data.

#### 4.3.1 Topic modelling

The entities that make up the topic modelling notion are “words,” “documents,” and “corpora.” As the basic unit of discrete data in a document, “word” is defined as vocabulary items that are indexed for each unique word in the document. “Document” is an arrangement of N words. A corpus is a collection of M documents, and its plural form, corpora, is corpora. While “subject” refers to the spread of a set vocabulary. Simply said, each document in the corpus has its own proportion of topics mentioned based on the terms it includes. Topic modelling is a prevalent technique for identifying hidden patterns or topics within a big corpus of text data. Particularly effective for discovering and analysing the underlying subjects inside social media data, such as tweets about supply chain firms and their emissions. Latent Dirichlet Allocation (LDA) is a generative probabilistic model that represents a document as a mixture of themes, where each topic is a probability distribution over a set of words. To do topic modeling, we first tokenized the tweets, removed stop words and other noise, and converted the text to lower case. The LDA model was constructed using the Gensim Python module. The model was trained on the whole corpus of tweets pertaining to emissions and supply chain firms, and we tested with various numbers of subjects to determine the ideal configuration. After training the LDA model, we extracted the most representative terms for each topic and labelled them based on their content. Then, we manually evaluated and analysed the labelled topics to determine the data’s significant themes and patterns. This strategy allowed us to obtain a deeper grasp of the Twitter debate surrounding emissions and supply chain firms and to select the most pertinent and significant subjects for our research. Overall, topic modelling proved to be a potent and effective technique for discovering hidden patterns and

topics within the enormous corpus of social media data that we collected, and it provided useful insights into the Twitter discourse around supply chain firms and their emissions.

### 4.3.2 Sentiment analysis using Naïve Bayes

Since Sentiment Analysis (SA) is a trending issue in the field of Natural Language Processing (NLP), there are numerous methods to determine the state of sentiment (positive or negative feeling) inside a text, paragraph, or the entire document. Sentiment Analysis follows a specific approach that begins with the collection and identification of data. Later, the necessary features will be extracted in preparation for the following stage, which is sentiment categorization. Ultimately, decision-making will occur at the phase of sentiment polarity and subjectivity. Sentiment analysis typically employs machine learning and lexicon-based methodologies (Medhat et al., 2014) to investigate the emotion contained within a document. According to Textblob documentation (Loria, 2020), Textblob uses the Nave Bayes (NB) classification model (refer to Fig. 2). The NB classifier was trained on NLTK (Natural Language ToolKit) to determine the sentiment of aggregated tweets.

Figure 2 shows the hierarchy of Naive Bayes in sentiment analysis within the broader field of machine learning. Sentiment analysis involves the use of computational tools and techniques to automatically classify text into different categories of sentiment, such as positive, negative, or neutral. One of the subcategories within sentiment analysis is lexicon-based algorithms, which use pre-defined dictionaries of words and their associated sentiment scores to classify text. Another subcategory is supervised learning, which involves training a machine learning model on labeled data to classify new, unlabeled data. Unsupervised learning, on the other hand, involves clustering or grouping data without any prior knowledge or labels. Probabilistic classifiers are a common approach used in sentiment analysis and machine learning more broadly. Naive Bayes is a type of probabilistic classifier that assumes independence between the features of the data. In the context of sentiment analysis, the features are the words in the text, and Naive Bayes calculates the probability of a given text belonging to a particular sentiment category based on the frequency of the words in that text. The hierarchy in Fig. 2 suggests that Naive Bayes is one of the methods used within the broader category of probabilistic classifiers for sentiment analysis.

NB is a probabilistic technique that uses Bayes’ theorem to determine sentiment distribution over the data. However, NB reduces all text to a bag of words, which disregards the placements of the words entirely.

$$P\left(\frac{label}{features}\right) = \frac{P(label) * P(features/label)}{P(features)}$$

where P(label) represents the prior probability of a label, P(features/label) represents the prior probability of a given feature set being classed as a label, and P(features) represents

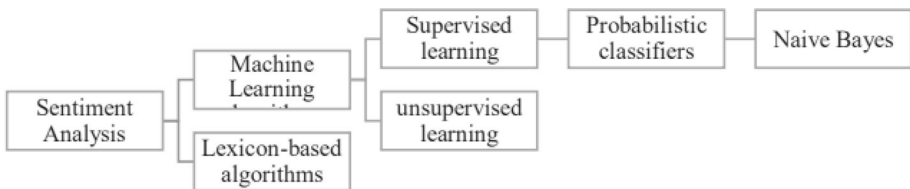


Fig. 2 Naïve Bayes Hierarchy (Source: Manguri et al., 2020)

the prior probability of a given feature set happening. We chose this approach because it is a well-established and widely used method for sentiment analysis (Alahmary & Al-Dossari, 2023; AminiMotlagh et al., 2022; Astuti et al., 2023), and it has been successfully applied to similar datasets in previous research. Furthermore, we wanted to ensure the efficiency and speed of the sentiment analysis process, given the large volume of tweets we collected. Overall, the Naïve Bayes algorithm was a suitable choice for our research objectives, and its use allowed us to efficiently and accurately analyze the sentiment of the tweets related to emissions and supply chain companies.

### 4.3.3 Emotions in text

To analyse the emotions based on text in cleaned tweets, we resorted to a new package known as text2emotions API to extract the emotions in the tweets.<sup>1</sup> Since text2emotion is a new technique, the extracted emotions were manually cross validated. In addition, we used the VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon, a sentiment analysis tool, to determine the sentiments expressed in tweets and cross-validated the outputs of text2emotion and VADER. The text2emotion with VADER's output. The text2emotion API generated several scores as follows

$$\{s_{t,e} \text{ for each } e \in E\}$$

where E is a set of basic emotions which can be represented as:

$$E = \{\text{set of basic emotions}\} = \{\text{angry, fear, happy, sad, surprise}\}$$

The emotion score for each set of emotion in set E is derived from the equation as follow:

$$\begin{aligned} \text{Score\_happy}_{x,d} &= \text{Happiness expressed in all tweets pc, dc, poc for tweet } X \text{ on day } d \\ &= \sum_{t \in T_{x,d,e=\text{happy}}} S_{t,e} \end{aligned}$$

where pc = pre-covid tweets.

dc = during covid tweets.

poc = post-covid tweets.

In a similar way, we calculated scores for anger, sad, surprise, and fear. The purpose of analysing emotions in text is to gain a deeper understanding of the sentiment and attitudes expressed in the tweets related to emissions and supply chain companies. By using both the text2emotion API and VADER lexicon, we were able to cross-validate the emotions extracted from the tweets and ensure the accuracy of our results. The emotion analysis provides a complementary perspective to the sentiment analysis conducted using the Naïve Bayes algorithm, allowing us to gain a more nuanced understanding of the emotional tone of the tweets. Overall, the emotion analysis technique adds value to our research by providing insights into the emotions expressed by users towards emissions and supply chain companies, which can be useful for companies and policymakers in addressing issues related to emissions and sustainability.

<sup>1</sup> <https://pypi.org/project/text2emotion/#description>.

## 5 Analysis and findings

### 5.1 Preliminary analysis

In this section, we present our findings. Table 1 shows the collection of tweets pre, during and post-COVID. The timeline for pre-COVID is from Jan 2019-Jan 2020 giving us 1,284,897 tweets, for during the COVID period, from Feb 2020-Dec 2021 giving us 815,367 tweets and post-COVID timeline is Jan 2022-June 2022 providing us with 115,141 tweets.

The total sample size is 2,215,678 tweets. In the next step, we filtered out the tweets by using keywords; “Emission”, “Emissions”, and “Pollution”. Now our sample size reduced to 3122 tweets out of our initial sample of 2,215,678 tweets. Table 2 shows the breakup of focussed tweets pre, during and post-COVID.

In Table 3, we presented the preliminary analysis of replies, retweets, likes and quotes on tweets for three phases of COVID. Twitter offers users a number of interaction functions in the current platform version (2021) so they can engage with tweets on their own. The four primary forms of engagement functionalities—like, retweet, quote tweet, and reply. These interaction strategies have various inputs and results. While quoting and commenting are considerably more informational and conversational because they allow users to convey unique ideas and content, liking and retweeting are relatively basic and easy engagement behaviours because they both lack additional original content. Regarding output, aside from liking, the other three forms of engagement behaviours (i.e., retweets, quote tweets, and replies) might result in new tweets that are tallied toward a user’s overall number of tweets posted (Fang et al., 2022).

In pre and during COVID timelines, the statistics related to replies, retweets, liked and quotes are similar, for e.g., Std. Dev of replies in pre-COVID is 24.90, and Std. Dev of replies during COVID is 28.36. The mean of retweets is 87.53, 90.24 and 34.41 for replies on tweets for pre, during and post-COVID respectively.

**Table 1** Collection of data

Period	Timeline	Total Tweets
Pre-COVID	Jan’2019–Jan’2020	1,284,897 tweets
COVID	Feb’2020–Dec’2021	815,367 tweets
Post COVID	Jan’2022–June’2022	115,141 tweets

\*Total Companies: -73

**Table 2** Focussed tweets

	Pre-COVID	COVID	Post COVID
Emission	797	683	73
Emissions	683	567	44
Pollution	186	71	18
Total	1666	1321	135

**Table 3** Preliminary analysis

	Count of tweets	Mean	Std. Dev	Min	Max
<i>Pre-COVID- Jan-2019–Jan-2020</i>					
Replies on Tweets	1666	6.48	24.90	0	504
Retweet on Tweets	1666	23.93	87.53	0	1560
Likes on Tweets	1666	240.71	1184.72	0	17,305
Quote on Tweets	1666	2.73	11.27	0	260
<i>COVID- Feb-2020–Dec-2021</i>					
Replies on Tweets	1321	10.83	28.36	0	504
Retweet on Tweets	1321	32.88	90.24	0	1320
Likes on Tweets	1321	334.75	1264.75	0	17,288
Quote on Tweets	1321	3.75	10.22	0	114
<i>Post-COVID- Jan-2022–June-2022</i>					
Replies on Tweets	135	13.33	23.07	0	159
Retweet on Tweets	135	22.17	34.41	0	145
Likes on Tweets	135	163.95	281.02	0	1061
Quote on Tweets	135	3.60	9.41	0	77

## 5.2 Content analysis

Figures 3, 4 and 5 shows the textual analysis of tweets for pre-, during and post-COVID tweets respectively. Figure 3 delineates the frequency distribution of pivotal phrases in tweets collected from the pre-COVID-19 timeline. Displayed on the x-axis are the specific terms under analysis, while the y-axis illustrates their corresponding frequency within the corpus. Intriguingly, the term 'emissions' stands out with a frequency exceeding 1,400 instances, highlighting its centrality in supply chain discussions related to decarbonization. The term 'CO,' cited around 600 times, suggests a focus on carbon monoxide or carbon-related topics. The prominence of 'consumption' and 'fuel,' with frequencies of approximately 550 and 430 respectively, possibly indicates concerns related to energy usage and its environmental implications. Lastly, the term 'greenhouse' occurs less frequently, with around 200 mentions, signaling that it might not be as primary a focus in pre-COVID discussions within supply chain contexts. The distribution of these terms' sheds light on what aspects of decarbonization were most discussed before the onset of the pandemic, thereby serving as a baseline for comparison with subsequent periods. Figure 4 provides a detailed analysis of phrase frequency for tweets during the COVID-19 timeline. While the general pattern is congruent with that observed in Fig. 3, there are some noteworthy deviations. For instance, the term "consumption" ascends to 5th place in frequency rankings, up from its 6th place standing in the pre-COVID-19 data. This shift could be indicative of growing awareness or concern about sustainable consumption practices during the pandemic. Interestingly, the term "greenhouse" declines in prominence, moving to the second-to-last position with approximately 180 mentions. This reduced frequency may suggest that greenhouse-related topics received lesser attention during the pandemic. Moreover, the term "carbon" gains traction, registering 300 mentions, a trend that could reflect increased focus on carbon management strategies amid the crisis.

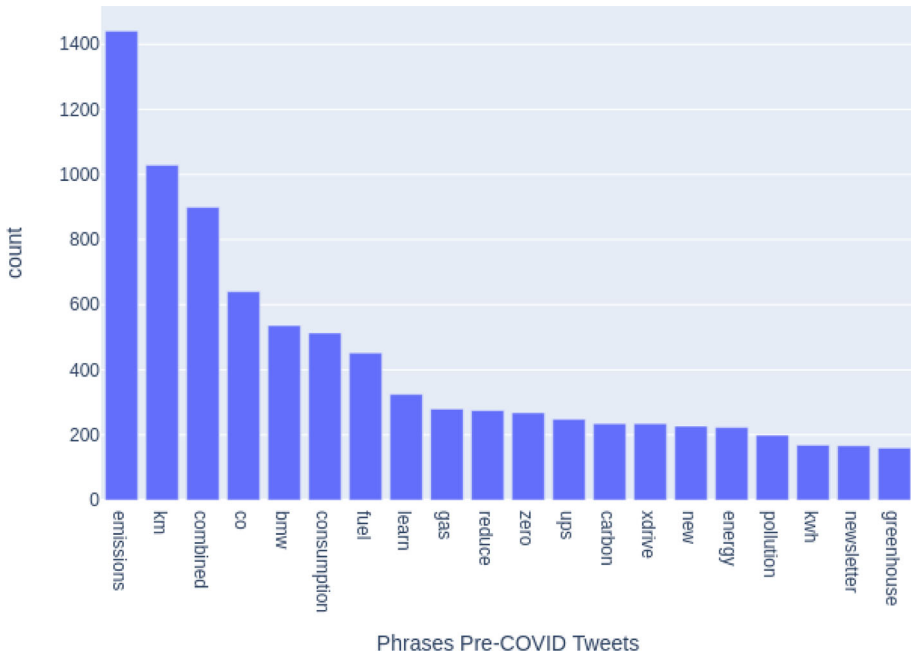


Fig. 3 Phrases pre-COVID Tweets

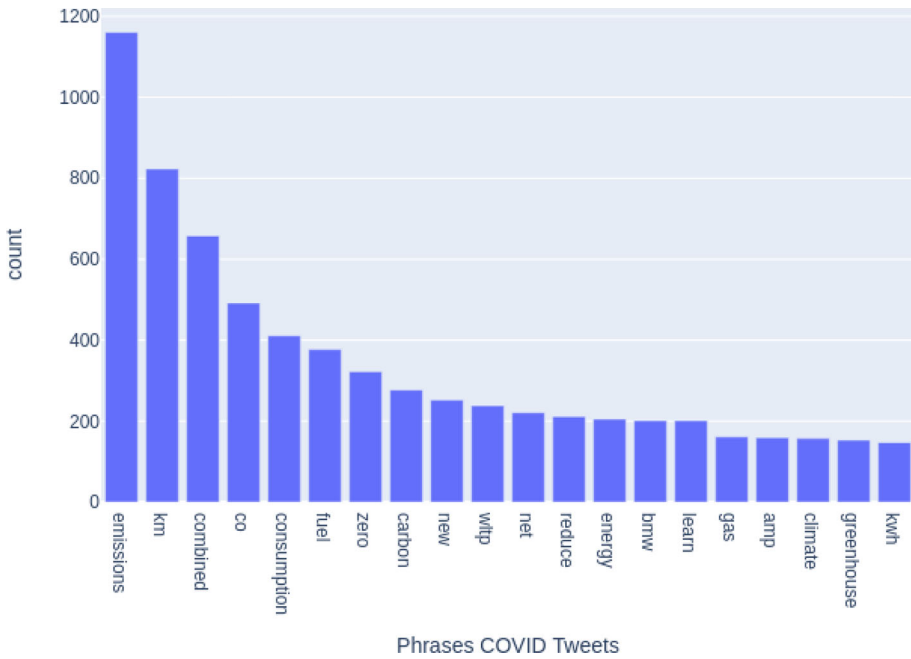
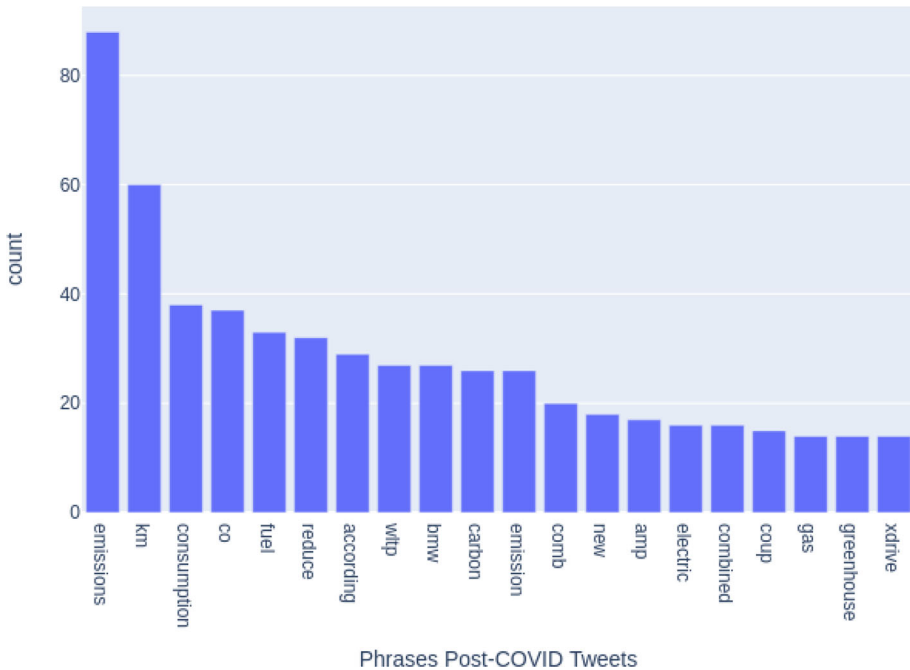


Fig. 4 Phrases COVID Tweets



**Fig. 5** Phrases post-COVID Tweets

Figures 6, 7 and 8 displays the word-cloud for the tweets for all the three phases of COVID-19.

Using a corpus of tweet tokens, dominating keywords can reveal the most popular subjects mentioned by users. A graphic representation of the frequency of an n-count of terms is called a word cloud. In contrast, smaller words are linked to a lower frequency. Bolder and bigger terms in a word cloud reflect keywords that are used more frequently by users. In these word-cloud, the more a specific word appears in a tweet, the bigger and bolder it appears in the word-cloud. During the phase of COVID, and post-COVID, the emission word is dominant followed by CO (carbon dioxide). In pre and during COVID timelines, the fuel consumption word also dominating the landscape of tweets. In pre-COVID and during COVID's word cloud, the dominant keywords are energy consumption, plastic pollution, gas emissions, zero emissions, air pollution, natural gas, whereas as post-COVID word cloud, help, global, electric, greenhouse gas, carbon emissions are the dominant keywords. In Figs. 9, 10, and 11, we presented the polarity analysis of tweets for three phases. In all three phases, the positive sentiment dominating the negative sentiments.

In Fig. 6, the words "emission" and "consumption" are the most prominent words in the word cloud, indicating that these are the most common topics discussed by the supply chain companies in their pre-COVID-19 tweets. The words "zero emission" and "energy consumption" are also included in the word cloud, suggesting that supply chain companies were already discussing and promoting zero-emission technologies and practices even before the COVID-19 pandemic. Finally, the words "climate change" and "combined" further emphasize the environmental concerns and efforts being discussed by these companies even prior to the pandemic. Overall, the word cloud for pre-COVID-19 tweets indicates that supply chain



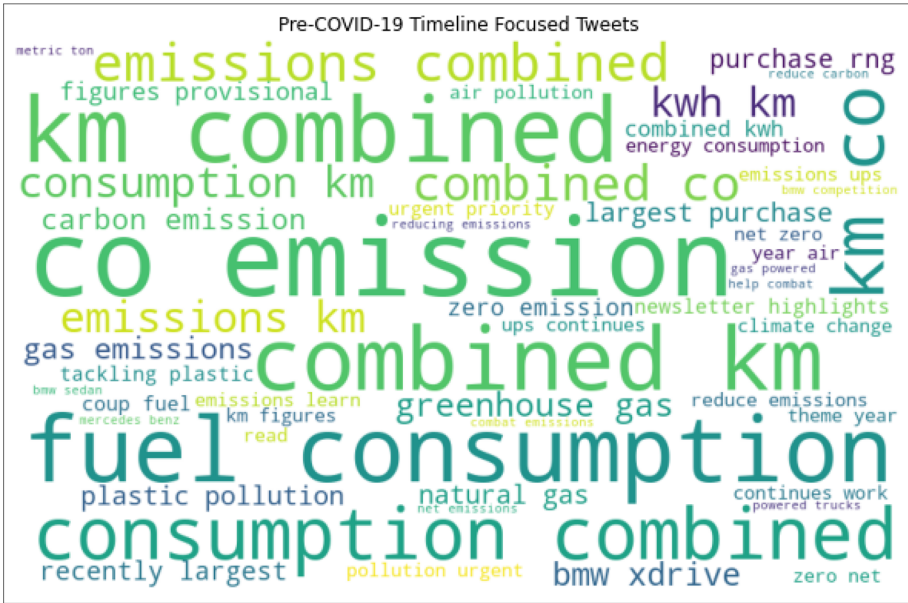


Fig. 6 Word cloud pre-COVID tweets

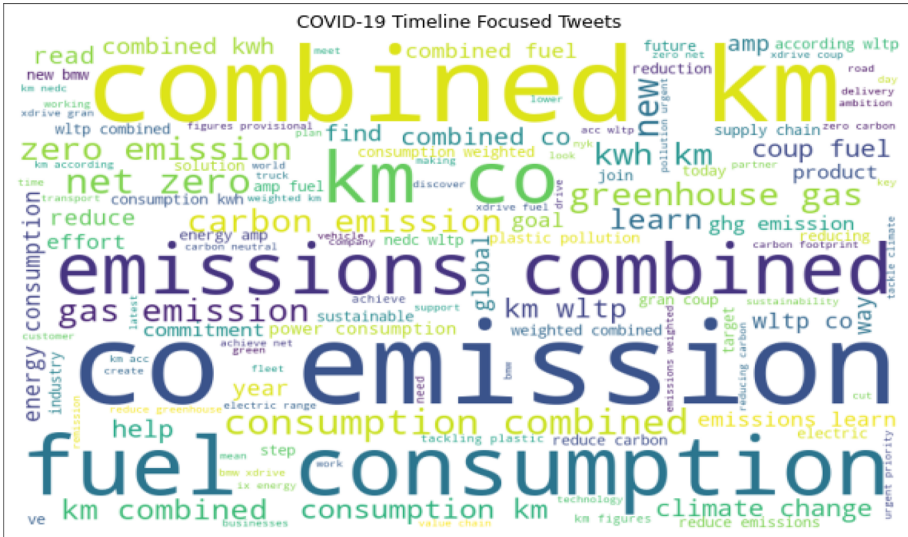


Fig. 7 Word cloud COVID tweets

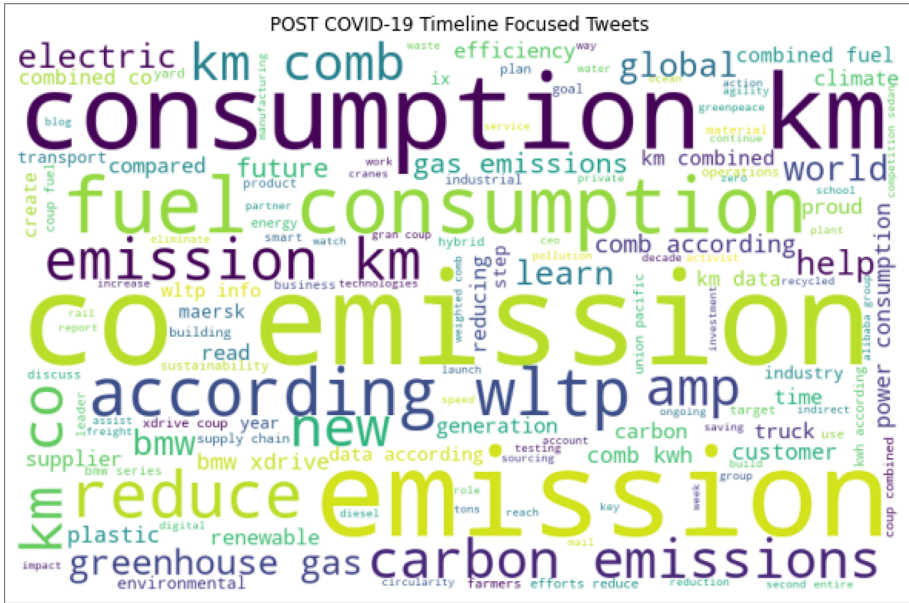


Fig. 8 Word cloud post-COVID tweets

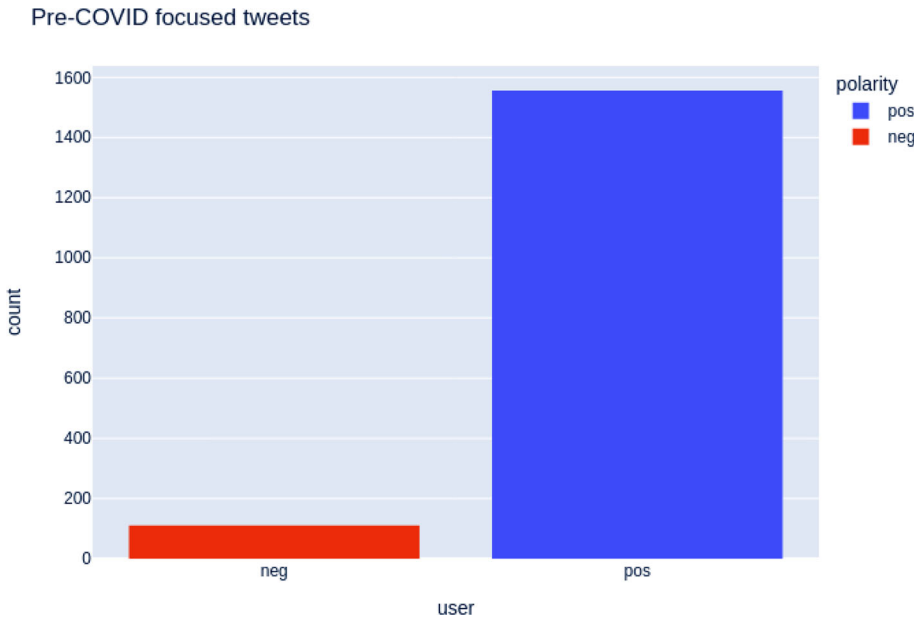


Fig. 9 Polarity of pre-COVID tweets

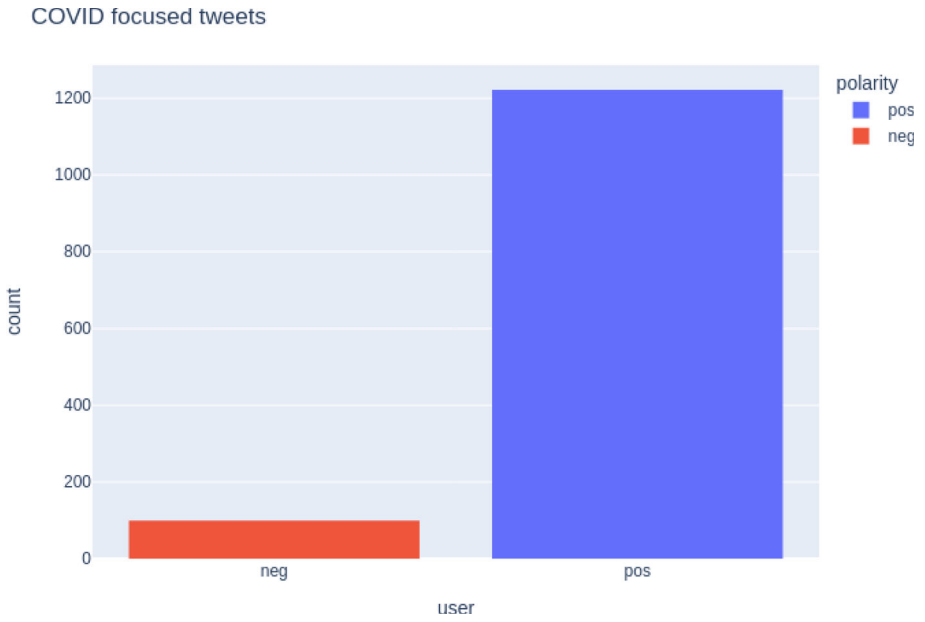


Fig. 10 Polarity of COVID-19 tweets

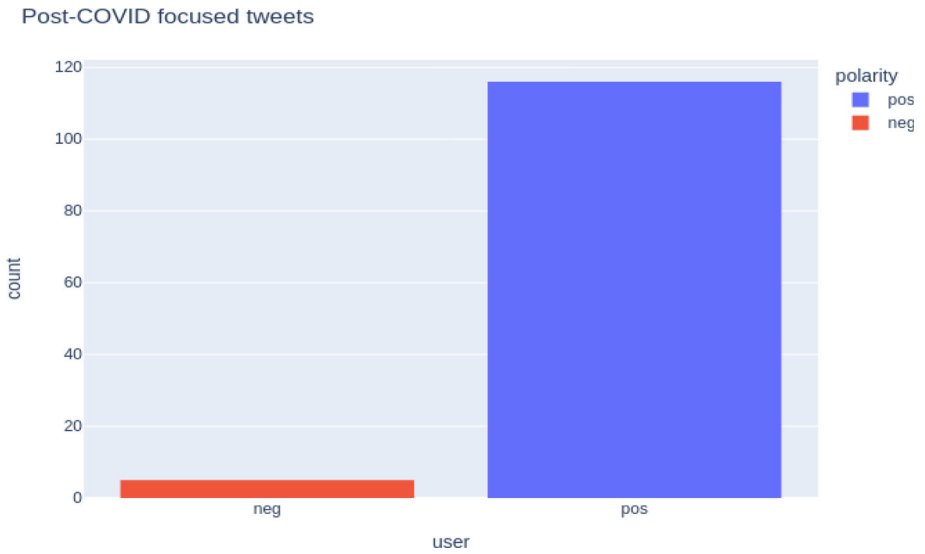


Fig. 11 Polarity of post-COVID tweets

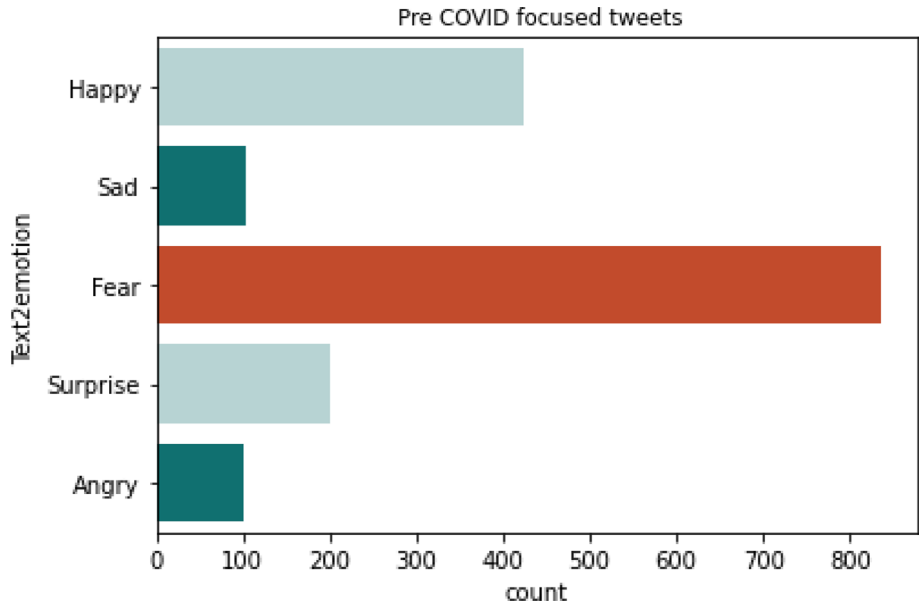
companies were already aware of the need for emission reduction and energy efficiency in their operations.

Figure 7 represents a word cloud generated from the tweets collected during the COVID-19 timeline, which were filtered on the basis of keywords mentioned in methodology. The word cloud provides a visual representation of the most frequently occurring words in the tweets, with the size of each word indicating its relative frequency in the dataset. As we can see, the word cloud includes a variety of keywords related to emissions and sustainability, such as emission, sustainable, carbon neutral, global, and carbon market. This suggests that supply chain companies were actively discussing and promoting sustainable practices during the COVID-19 pandemic. Additionally, the presence of words such as km and fuel suggest that companies were also discussing the reduction of fuel consumption and emissions through the use of more sustainable transportation methods. Overall, this word cloud provides insight into the topics that were most commonly discussed on Twitter by supply chain companies during the COVID-19 pandemic related to emissions, sustainability, and reducing carbon footprint.

Figure 8 presents the word cloud generated for the post-COVID-19 tweets collected from supply chain companies, after filtering them based on the presence of keywords "emission", "emissions", and "pollution". A word cloud is a visualization tool that displays the most frequently used words in a text corpus, with the size of each word indicating its frequency of occurrence. In this word cloud, we can observe that the most prominent words are "consumption", "electric", "emission", "greenhouse", "carbon", "gas", "co", and "global". These words are indicative of the topics and themes that were most commonly discussed in the post-COVID-19 tweets related to emissions and pollution by the supply chain companies. The presence of words like "consumption" and "electric" suggests that there is an increasing focus on reducing emissions through the adoption of electric vehicles and other forms of sustainable transportation. The words "emission", "greenhouse", "carbon", and "gas" indicate that the companies are actively discussing ways to reduce their carbon footprint and greenhouse gas emissions. Furthermore, the appearance of the word "global" in the word cloud indicates that supply chain companies are acknowledging the global impact of their operations and are taking steps to mitigate their environmental impact. Overall, this word cloud provides an insight into the topics and themes being discussed by supply chain companies in relation to emissions and pollution in the post-COVID-19 era.

The word cloud for pre-COVID tweets shows prominent words such as "emission," "consumption," and "climate change," indicating the companies' concern for environmental sustainability before the pandemic. The word cloud for COVID tweets highlights words such as "sustainable," "carbon-neutral," and "reduce," which suggests the companies' focus on reducing emissions during the pandemic, likely driven by societal and economic pressures. Finally, the word cloud for post-COVID tweets shows a shift towards "electric" and "zero emission," indicating the companies' focus on transitioning to more sustainable and cleaner energy sources. Overall, these word clouds reveal the changing priorities of supply chain companies over time, reflecting their response to environmental and social pressures.

Figure 9 shows polarity analysis of pre-COVID tweets related to emissions and supply chain companies. The figure shows that there were 1500 tweets with a positive polarity and 150 tweets with a negative polarity. Polarity analysis is a common method used in sentiment analysis to determine the overall positive or negative sentiment of a given text. In this case, the polarity analysis indicates that the majority of pre-COVID tweets related to emissions and supply chain companies were positive in sentiment, with only a small minority being negative. This information can be useful in understanding the general sentiment and attitudes of users towards emissions and supply chain companies before the COVID-19 pandemic.



**Fig. 12** Pre-COVID emotion analysis of tweets

Furthermore, the results of the polarity analysis may also suggest that pre-COVID users were generally supportive of efforts related to emissions reduction and sustainability within the supply chain industry. The relatively low number of negative tweets may also indicate that there was less public scrutiny or negative attention given to supply chain companies regarding their emissions practices prior to the pandemic.

We also plotted the results of emotion analysis of tweets for all three phases by using a text2Emotion package in python. It processes any textual material, detects the emotion it contains, and outputs the results as a dictionary in form of five basic emotion categories such as Angry, Surprise, Sad, Happy and Fear. Figures 12, 13 and 14 shows the five emotions related to the tweets. In all three phases, fear is the dominant emotion followed by happy.

Based on the analysis presented in Fig. 12, it appears that the majority of tweets related to emissions and supply chain companies before the COVID-19 pandemic were categorized as "fear" with approximately 800 tweets falling into this category. "Happy" was the second most frequent emotion with approximately 400 tweets, followed by "surprise" with around 200 tweets. "Sad" and "angry" were the least common emotions, each with around 100 tweets. These results can provide insight into the general emotional tone of tweets related to emissions and supply chain companies before the COVID-19 pandemic. Nonetheless, the results suggest that fear and happiness were the predominant emotions expressed in tweets related to emissions and supply chain companies before the COVID-19 pandemic.

Similarly, Fig. 13 shows emotion analysis of tweets during the COVID-19 pandemic related to emissions and supply chain companies. The majority of tweets categorized under this analysis were classified as "fear," with approximately 520 tweets falling under this category. "Happy" was the second most frequent emotion expressed in tweets, with around 340 tweets, followed by "surprise" with approximately 180 tweets. "Sad" and "angry" were the least common emotions, each with approximately 140 tweets. These results suggest that, during the COVID-19 pandemic, fear continued to be the predominant emotion expressed in

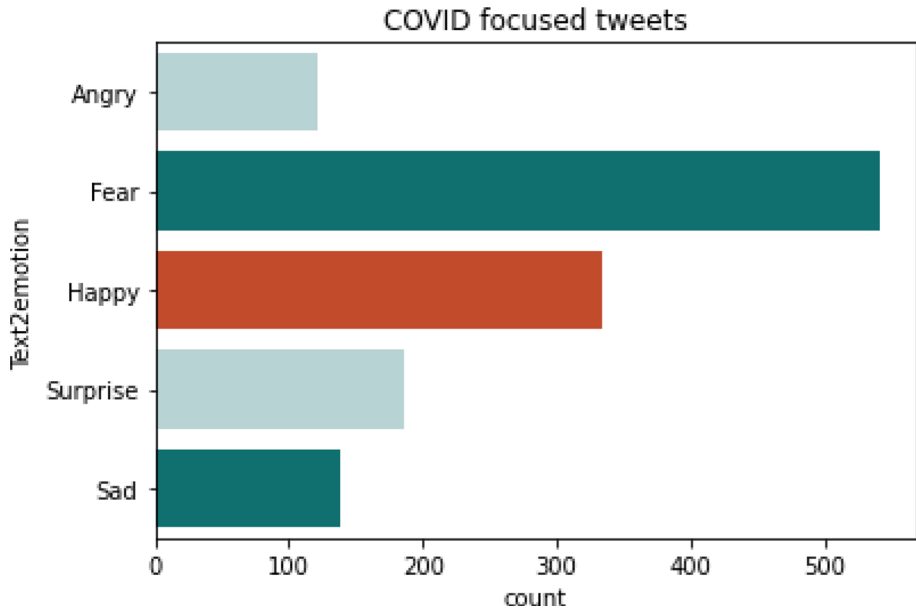


Fig. 13 COVID emotion analysis of tweets

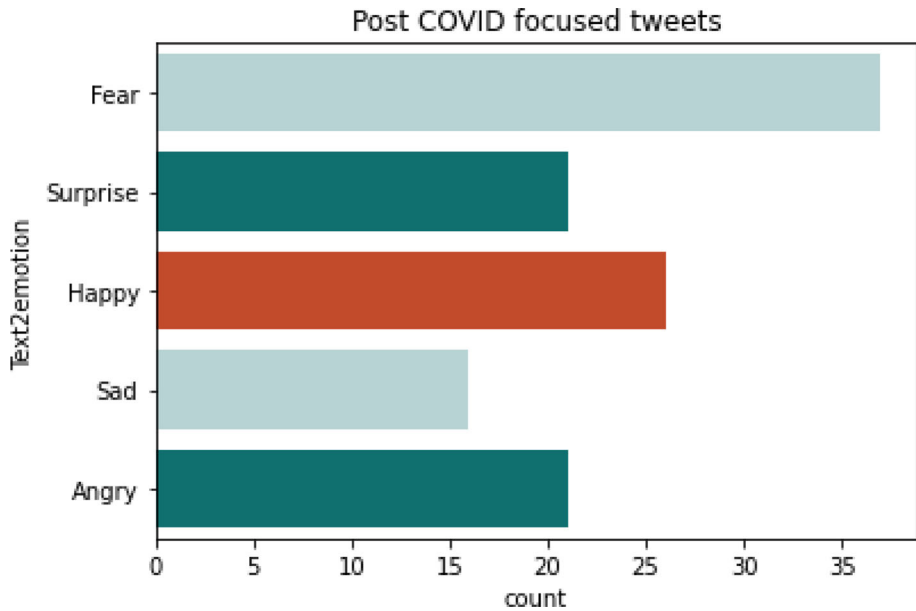


Fig. 14 Post COVID emotion analysis of tweets

tweets related to emissions and supply chain companies. However, it is interesting to note that the number of tweets expressing "happy" emotions decreased significantly compared to the pre-COVID-19 period. Overall, this analysis highlights the potential impact of the COVID-19 pandemic on the emotional tone of tweets related to emissions and supply chain companies.

Based on the analysis presented in Fig. 14, it appears that the emotional tone of tweets related to emissions and supply chain companies has shifted after the COVID-19 pandemic. The results indicate that "fear" remains the most common emotion, but with a significantly lower count of only 36 tweets. "Happy" is the second most frequent emotion with 25 tweets, followed by "surprise" with around 20 tweets. "Angry" and "sad" were the least common emotions, each with around 21 and 15 tweets, respectively. These results suggest that the emotional tone of tweets related to emissions and supply chain companies has shifted to a more positive direction after the COVID-19 pandemic, with a lower emphasis on negative emotions such as fear and anger.

In Figs. 15, 16, and 17, we plotted the subjectivity and polarity scores through a scatter plot for all the phases. The classification of sentences as subjective judgments or objective facts is subjectivity analysis. Therefore, using subjectivity analysis, we separate subjective statements from objective sentences for a set of messages in a dataset. Our dataset has been subjectivity-analysed using Python TextBlob for Natural Language Processing (NLP). Python TextBlob, a well-known NLP programme that is free to use, can be employed to do out text subjectivity analysis. The textblob library, which employs a built-in model to determine the subjectivity value, was used to calculate the subjectivity of each tweet. The range of

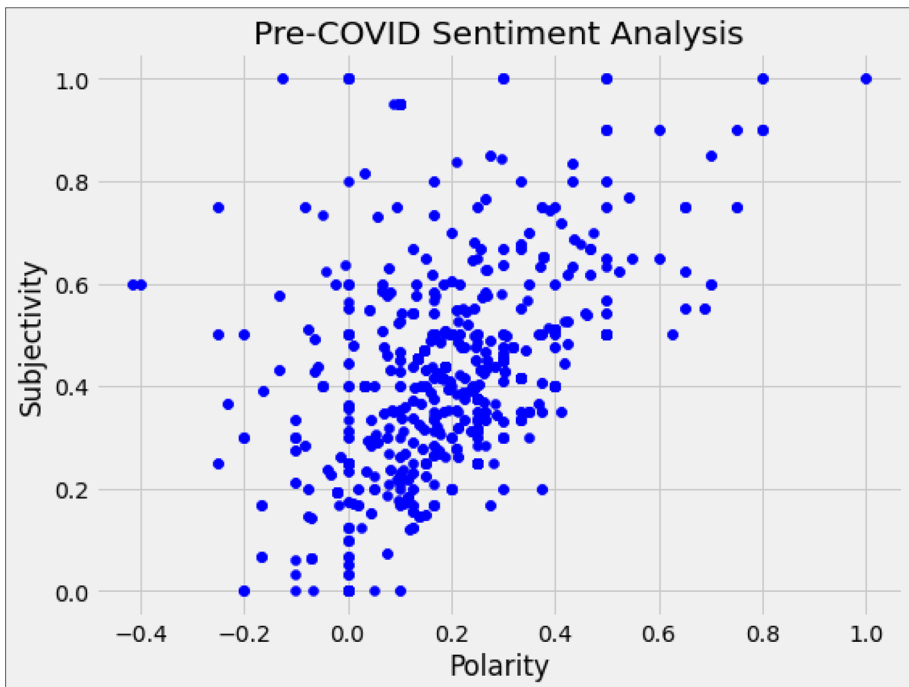


Fig. 15 Scatter plot of subjectivity and polarity pre-COVID

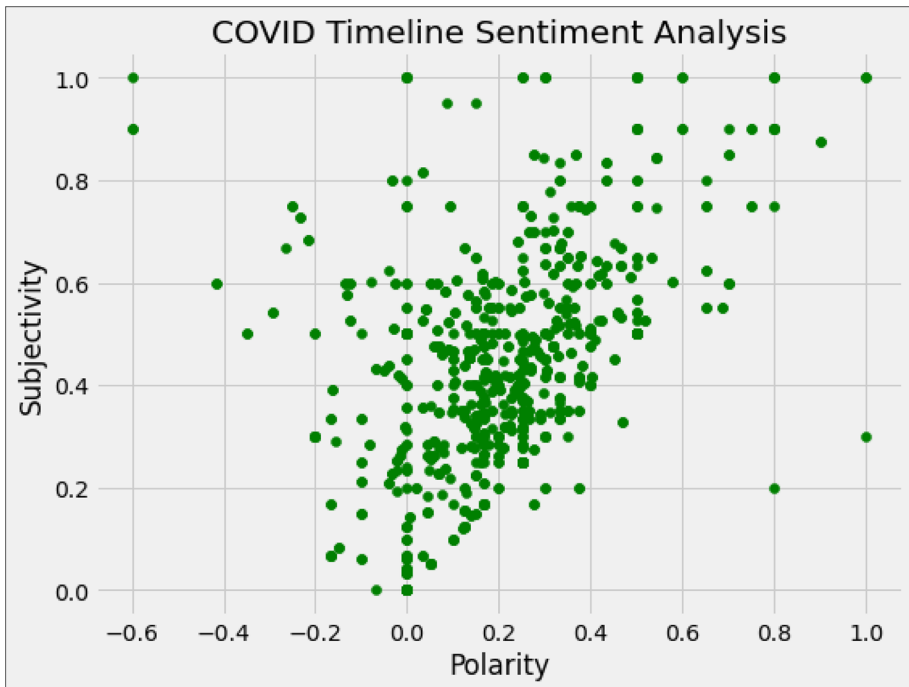


Fig. 16 Scatter plot of subjectivity and polarity COVID

subjectivity values are shown in Table 1. An objective text is indicated by a number close to 0, whereas a highly subjective text is indicated by a value close to 1. After determining if a statement is opinionated or not through subjectivity analysis, we analyse the text for polarity to determine whether it communicates a positive or negative opinion. The goal of polarity analysis is to ascertain the text writer's emotional stance toward the subject being discussed. As Twitter becomes more and more popular as a communication tool, interest in using its data to gauge public sentiment toward a subject or well-known figures has also increased. We can get that knowledge by quantifying the sentiment of text using polarity analysis. Thus, the text can be categorised as either negative, positive, or neutral. There are numerous technologies available for brief text sentiment analysis. In all the phases of COVID, we can observe a positive linear relationship between subjectivity and polarity of the tweets.

Tables 4, 5 and 6 show the ten most prevalent topics in our sample dataset for all the three phases. We also plotted the frequency of these topics by count for all three phases in Figs. 18, 19 and 20. Topic modelling is an unsupervised classification of documents, akin to clustering of numerical data. We used Latent Dirichlet Allocation (LDA) technique. It maximises the likelihood that a word belongs to a given topic. It is predicated on the idea that a corpus of text can be represented by a distribution of topics, with each subject being characterised by a distribution of words. The probability of each word's appearance in a document is then used to assign it to a topic, creating a topic mixture for the content. This method has been used to analyse online support group posts to find cogent issues in social science contexts. In Table 4, which show the topics for pre-COVID phases, the top topics with keywords are emissions, carbon, progress at place 0, the topic 2 comprises of words such as emissions, gas, carbon,



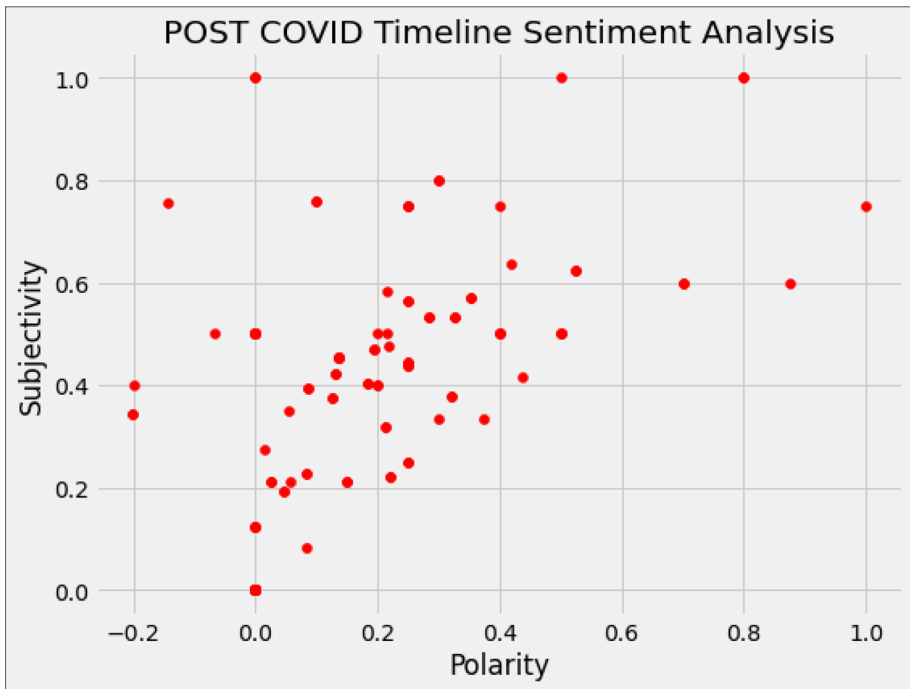


Fig. 17 Scatter plot of subjectivity and polarity post-COVID

Table 4 Topic modelling pre-COVID tweets

Pre-COVID Focused Tweets

Topic S.No	Topic
0	Emissions zero carbon net committed nestl progress ll
1	Emissions gas carbon reducing greenhouse zero new co
2	Emissions gas reduces ups learn natural net carbon
3	km combined bmw consumption fuel emissions co xdrive
4	km combined co emissions kwh consumption fuel energy
5	Emissions today pollution work air sustainability year find
6	Emissions learn reducing goals metric work tons recently
7	Emissions reduce use co greenhouse environmental plastics sustainability
8	Plastic emission pollution free help waste option problem
9	Waste plastic pollution tackling hi priority urgent working

**Table 5** COVID topic modelling

COVID Focused Tweets	
Topic S.No	Topic
0	Emissions reduce carbon greenhouse help gas impact co
1	Emissions zero carbon net climate global change read
2	km combined emissions co consumption fuel bmw coup
3	3 emissions carbon emission zero gas greenhouse new fuel
4	Emissions zero learn net hybrid reduce co change
5	km wltp combined consumption emissions co weighted fuel
6	Emissions zero net learn gas greenhouse climate carbon
7	km emissions combined consumption pollution plastic co kwh
8	Emissions zero electric climate remission rail trucks goal
9	Emissions carbon reduce amp energy fuel gas reducing

**Table 6** Post-COVID topic modelling

Post- COVID Focused Tweets	
Topic S.No	Topic
0	Ups newsletter emissions subscribe learn largest purchase recently
1	Emissions gas zero natural net learn reduce climate
2	km combined bmw consumption emissions co fuel xdrive
3	Emissions co km kombiniert mercedes vehicles electric benz
4	Emissions reduce gas new renewable greenhouse free ground
5	Emissions ghg rail reduces freight carbon zero million
6	Emissions co learn carbon zero reduction read chain
7	Emissions carbon learn reducing sustainable amp committed reduce
8	Emissions carbon help new zero co address energy
9	Plastic pollution tackling urgent waste priority hi packaging

Pre-COVID focused Tweets Topic Modelling

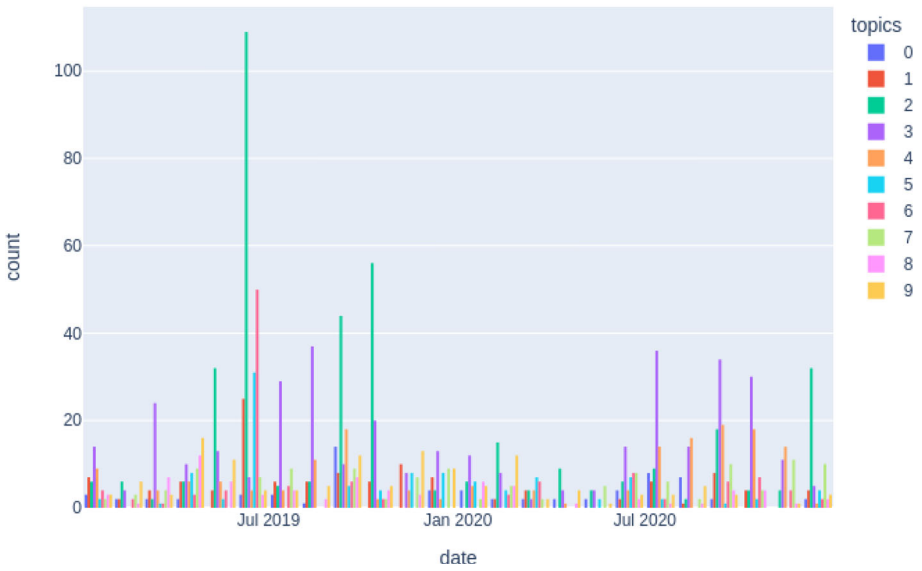


Fig. 18 Pre-COVID topics frequency

COVID focused Tweets Topic Modelling

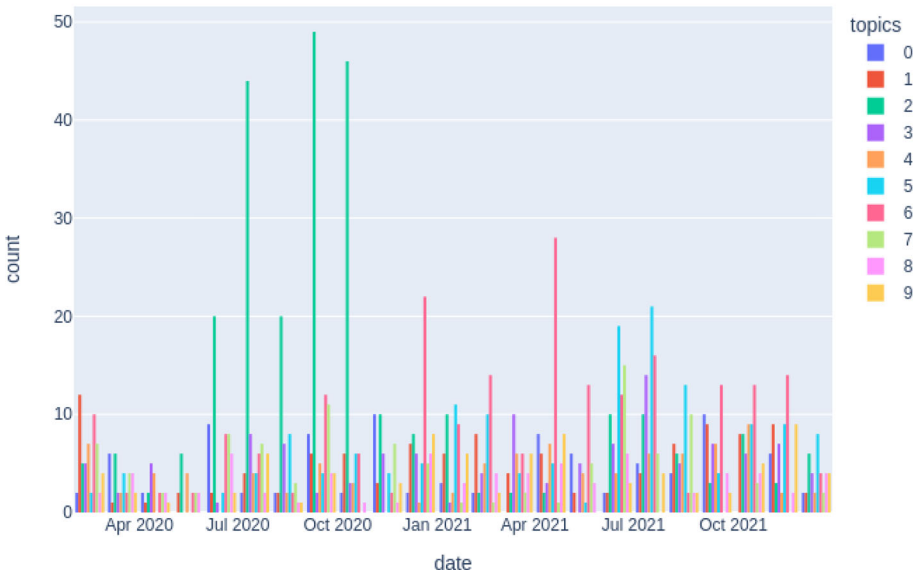
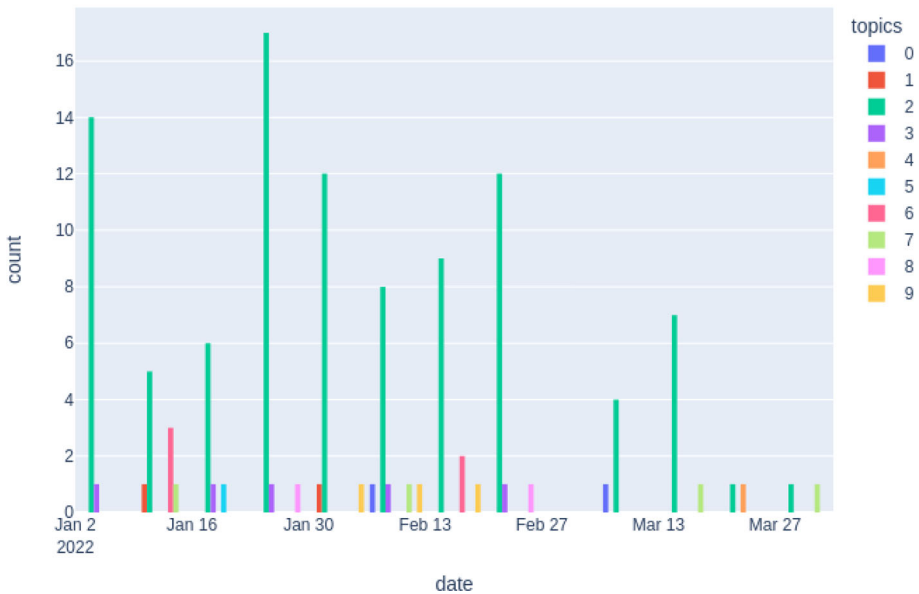


Fig. 19 COVID topic frequency

## POST COVID focused Tweets Topic Modelling



**Fig. 20** Post-COVID topic frequency

reducing, greenhouse topped with count of more than 100 tweets during pre-COVID in July 2019. The frequency of rest of the topics varied from frequency of 40 to 20.

During COVID phase, topic 2 dominates with a frequency of 40 tweets at the start of October 2020 followed by 30 tweets at the end of October 2020. From Jan 2021 to April 2021, topic 1 with keywords like emissions, subscriber, learn, largest started to gain traction with a frequency of 24 tweets in Jan 2021 and reaches to 28 tweets during April 2021.

Table 6 show the post-COVID topic modelling with topic 2 dominating till the end of our sample period with keywords such as combined, consumption, fuel, emissions followed topic 1 with keywords such as newsletter, emission, subscribe, purchase etc.

## 6 Discussion and conclusion

This study makes a modest attempt to understand key discussions concerning decarbonization in supply chain through the lens of twitter data. By sourcing tweets from firms carefully selected Forbes 2000 companies and Gartner Top 50 supply chain companies, we carried out a step-by-step procedure to address our research questions with COVID-19 as the background context since reducing pollution seems to be secondary towards resilience in the recovery phase of the pandemic. With more than 22,00,000 tweets spanning more than three years, we carried out our preliminary analysis to understand how decarbonization is discussed during pandemic times. Surprisingly, although predictable, tweets concerning pollution and related keywords such as emissions was considerably low in proportion compared to the master list of tweets that we gathered. A careful reflection of the proportion points out that decarbonization

related discussions accounted for roughly 0.13% during pre-pandemic, 0.16% during pandemic and 0.11% post pandemic. This is counter-intuitive as common understanding suggest that pollution levels decreased during pandemic times. Although the proportion is too low to make an impact, it is unfortunate and therefore significant for firms to focus more towards emissions and pollution if nations as a whole wish to aim for carbon neutrality by 2050. Moving ahead, we carried out our content analysis across each timeline. One of encouraging signs of decarbonization is that while we analysed phrases across timelines, we found that the count of words such as “consumption” and “fuel” has gone up in terms of ranking from pre-covid to post covid times. Since our refined sample only contains tweets involving pollution and emissions, discussion on consumption and fuel certainly gives us a positive indication towards sustainable consumption and the need for cleaner fuels across supply chains. Even if we refer to the word cloud, we find emphasis on greenhouse gas emissions, fuel consumption, net zero, climate change among others. However, since the proportion regarding supply chain issues and emissions are very low, overall importance of such keywords remains questionable at the firm level. Further, with twitter being so proactive on important discussions and announcement, top supply chain firms should be more vocal towards participating in discussions concerning decarbonization or at least treat such social media platforms as a mode of advertisement towards their efforts towards environmental sustainability. Moving towards emotion analysis, the inclination towards “fear” emotion raises important questions concerning decarbonization or otherwise especially during pre-covid times. For instance, if the emotion on tweets concerning emissions and pollution is more towards fear across timelines, then the question arises on the reasons why such leading supply chain players are not participating in pollution mitigating strategies in social media platforms such as Twitter. We believe that just reporting such initiatives in sustainability reports will neither influence other actors within the supply chain nor help tweet emotions to move from “fear” to other positive emotions such as “happy” or “surprise”. The topic modelling results identified important topics leading to decarbonization efforts. For instance, plastic pollution mitigation efforts, use of renewable energy use in supply chains, greenhouse gas emissions in freight transport etc. are all critical issues concerning greening in supply chains.

## 7 Research implications

We posit certain key implications to research concerning twitter, supply chain and decarbonization. We extend the discussion of Chae (2015) concerning supply chain issues and twitter data and explain how there is a need to focus on specific environmental challenges within supply chains in addition to discussing environmental standards. Unfortunately, while Chae (2015) emphasis the need for stakeholder engagement and communication, our analysis did not find any focused involvement of stakeholder engagement initiatives towards decarbonization. Our results, however, point out that emission related discussion remains a fear emotion across pandemic timelines and therefore the impact of pandemic does not alter the importance of emission concerns within supply chain. This points out the relevance of Sharma et al. (2020) involving twitter data on the deployment of technology in not only improving supply chain resilience but also to reduce use of fuels that are causing harm to the decarbonization efforts. Further, linking to the discussion on environmental pollution and COVID-19 (Marwah & Agrawala, 2022; Keshtkar et al., 2022; Yang et al., 2021), we neither found evidence of decarbonization efforts in the post pandemic recovery phase nor found any

linkages on the discussion on pollution towards potential strategies towards reducing pollution in the long run. This may be because of two reasons: First, due to surge of economic activity in the recovery stage, firms felt the need to discuss the challenges of the recovery phase rather than focus on the increase in pollution levels due to increase in business activities. Second, firms were more focused towards making supply chain more resilient in the wake of similar pandemic in near future. However, this remains a concern for leading supply chain firms and environmentalists on the use and engagement of social media platforms towards highlighting the importance of decarbonization activities at various stages within the supply chain.

## 7.1 Managerial and policy implications

The managerial implications for supply chain decarbonization are numerous and varied, as the process of reducing a company's carbon footprint can involve changes across multiple areas of the organization. Some of the key managerial implications include: Commitment from top management: The commitment of senior management is critical to achieving supply chain decarbonization goals. Managers must be willing to invest resources, set targets, and make decisions that prioritize sustainability. Collaboration with suppliers: Supply chain decarbonization is often a collaborative effort involving multiple parties, including suppliers. Managers need to work with their suppliers to encourage sustainable practices and identify areas where emissions can be reduced. Integration of sustainability into procurement processes: Managers should consider sustainability factors when making procurement decisions, including selecting suppliers with strong environmental performance and considering the environmental impact of materials and products. Data collection and analysis: Effective supply chain decarbonization requires accurate and reliable data on emissions and other environmental impacts. Managers need to establish systems for collecting and analysing this data to inform decision-making and track progress towards sustainability goals. Communication and transparency: Managers should communicate their supply chain decarbonization efforts to stakeholders, including customers, investors, and employees. This communication can build support for sustainability initiatives and demonstrate the company's commitment to reducing its environmental impact. Overall, the managerial implications of supply chain decarbonization require a holistic approach to sustainability, with a focus on collaboration, data analysis, and effective communication. By taking these steps, managers can make progress towards reducing their company's carbon footprint while maintaining operational efficiency and meeting customer needs.

Overall, the future of decarbonization policies will likely involve a comprehensive and integrated approach that addresses various sectors of the economy and considers social, economic, and environmental factors. As the urgency of addressing climate change continues to grow, governments and stakeholders will need to work collaboratively to ensure a sustainable and low-carbon future. The commitment of political leaders and governments to prioritize decarbonization will have a significant impact on the pace and effectiveness of policy implementation. Public awareness and support for decarbonization will influence policy decisions. Governments may focus on educating the public about the benefits of sustainable living and encouraging behavioural changes that reduce carbon footprints. Policies promoting circular economy principles, which emphasize reducing, reusing, and recycling resources, can contribute to lower carbon emissions. Climate change is a global issue, and international collaboration is crucial. Agreements like the Paris Agreement provide a framework for countries to work together to limit global warming. Continued efforts to strengthen and implement such

agreements will be important. Continued investment in research and innovation is essential to discover new technologies and strategies for decarbonization.

## 7.2 Limitations and future scope

Our paper has several limitations which allows other researchers for further extension. First, since it was a tweets-based study, our results do not reflect the potential factors behind the outcome. For instance, individual companies have different approaches to emission mitigation and respective carbonization issues may be unique to a certain firm or industry. The tweets in our sample do not effectively reflect what are those factors or steps towards decarbonization within the industry or firm. Future research could specifically conduct sector specific work on decarbonization and ascertain how carbon emissions was better managed while firms recovered from the global pandemic. Second, unlike interview-based transcripts, tweets did not provide sufficient content on the critical issues leading to decarbonization. As a result, content analysis by coding broad themes was not possible in this work. Based on sector specific sentiments, broad themes could be generated about the issues concerning decarbonization by conducting interviews of top management professionals across geographies within a specific sector. This could complement our earlier recommendation for future work. The discussion is expected to bring in competing aspects of technology, cost, innovation, emissions in the backdrop towards the need for more resilient supply chains. Third, certain observation and results remain unclear. For instance, what could be the reasons fear emotions remained dominant across timelines especially pre-covid times. Since, our sample reflects emissions and pollution, it has to be further studies whether supply chains, irrespective of the global pandemic, are sceptical towards their decarbonization efforts and whether different stakeholders believe such efforts are in the right direction or not.

**Funding** Open access funding provided by Széchenyi István University (SZE).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alahmary, R., & Al-Dossari, H. (2023). A semiautomatic annotation approach for sentiment analysis. *Journal of Information Science*, 49(2), 398–410. <https://doi.org/10.1177/01655515211006594>
- AminiMotlagh, M., Shahhoseini, H., & Fatehi, N. (2022). A reliable sentiment analysis for classification of tweets in social networks. *Social Network Analysis and Mining*, 13(1), 7. <https://doi.org/10.1007/s13278-022-00998-2>
- Appiah, M., Li, M., Naeem, M. A., & Karim, S. (2023). Greening the globe: Uncovering the impact of environmental policy, renewable energy, and innovation on ecological footprint. *Technological Forecasting and Social Change*, 192, 122561.
- Astuti, I. F., Widagdo, P. P., Tanro, M. L. R., Cahyadi, D., & Suntura, A. A. (2023). Sentiment analysis on land and forest fire management in Twitter using Naïve Bayes method. *AIP Conference Proceedings*. <https://doi.org/10.1063/5.0110588>

- Chae, B. K. (2015). Insights from hashtag# supply chain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, *165*, 247–259.
- Chau, M., & Xu, J. (2012). Business intelligence in blogs: Understanding consumer interactions and communities. *MIS Quarterly*, *36*(4), 1189–1216. <https://doi.org/10.2307/41703504>
- Chon, M. G., & Kim, S. (2022). Dealing with the COVID-19 crisis: Theoretical application of social media analytics in government crisis management. *Public Relations Review*, *48*(3), 102201.
- Chung, S. H., Weaver, R. D., & Friesz, T. L. (2013). Strategic response to pollution taxes in supply chain networks: Dynamic, spatial, and organizational dimensions. *European Journal of Operational Research*, *231*(2), 314–327.
- Dai, R., Zhang, J., & Tang, W. (2017). Cartelization or Cost-sharing? Comparison of cooperation modes in a green supply chain. *Journal of Cleaner Production*, *156*, 159–173.
- Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. *California Management Review*, *54*(1), 64–87.
- Deloitte (2021). What is decarbonisation? The importance of decarbonisation explained. Accessed on August 30, 2022. Retrieved from <https://www2.deloitte.com/nl/nl/pages/energy-resources-industrials/articles/what-is-decarbonisation.html>
- Dutheil, F., Baker, J. S., & Navel, V. (2021). Air pollution in post-COVID-19 world: The final countdown of modern civilization? *Environmental Science and Pollution Research*, *28*(33), 46079–46081.
- El Ouardighi, F., Sim, J. E., & Kim, B. (2016). Pollution accumulation and abatement policy in a supply chain. *European Journal of Operational Research*, *248*(3), 982–996.
- Ernst, & Young (2022). How Europe's post-COVID-19 economy and environment can both prosper. Accessed on August 30, 2022. Retrieved from [https://www.ey.com/en\\_gl/eu-institutions/how-europe-s-post-covid-19-economy-and-environment-can-both-prosper](https://www.ey.com/en_gl/eu-institutions/how-europe-s-post-covid-19-economy-and-environment-can-both-prosper)
- Feng, Y., Lai, K. H., & Zhu, Q. (2022). Green supply chain innovation: Emergence, adoption, and challenges. *International Journal of Production Economics*, *248*, 108497.
- Georgi, C., Darkow, I. L., & Kotzab, H. (2010). The intellectual foundation of the journal of business logistics and its evolution between 1978 and 2007. *Journal of Business Logistics*, *31*(2), 63–109. <https://doi.org/10.1002/j.2158-1592.2010.tb00143.x>
- Janjua, N. K., Nawaz, F., & Prior, D. D. (2021). A fuzzy supply chain risk assessment approach using real-time disruption event data from Twitter. *Enterprise Information Systems*, *17*, 1959652.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68.
- Karim, S., Naeem, M. A., Tiwari, A. K., & Ashraf, S. (2023). Examining the avenues of sustainability in resources and digital blockchains backed currencies: evidence from energy metals and cryptocurrencies. *Annals of Operations Research*, 1–18.
- Kephart, J. L., Avila-Palencia, I., Bilal, U., Gouveia, N., Caiaffa, W. T., & Diez Roux, A. V. (2021). COVID-19, ambient air pollution, and environmental health inequities in Latin American cities. *Journal of Urban Health*, *98*(3), 428–432.
- Keshkar, M., Heidari, H., Moazzeni, N., & Azadi, H. (2022). Analysis of changes in air pollution quality and impact of COVID-19 on environmental health in Iran: Application of interpolation models and spatial autocorrelation. *Environmental Science and Pollution Research*, *29*(25), 38505–38526.
- Kouhizadeh, M., & Sarkis, J. (2018). Blockchain practices, potentials, and perspectives in greening supply chains. *Sustainability*, *10*(10), 3652.
- KPMG (2021). The decarbonisation journey: Five pillars to achieving net zero. Accessed on August 30, 2022. Retrieved from <https://home.kpmg/ua/en/home/insights/2021/09/five-pillars-of-net-zero.html>
- Kraines, S. (2002). Pollution and cost in the coke-making supply chain in Shanxi province, China: Applying an integrated system model to siting and transportation trade-offs. *Journal of Industrial Ecology*, *6*(3–4), 161–184.
- Loria, S. (2020). TextBlob: Simplified text processing. *Machine Learning Research*, *3*, 993–1022.
- Manguri, K. H., Ramadhan, R. N., & Amin, P. R. M. (2020). Twitter sentiment analysis on worldwide COVID-19 outbreaks. *Kurdistan Journal of Applied Research*. <https://doi.org/10.24017/covid.8>
- Marwah, M., & Agrawala, P. K. (2022). COVID-19 lockdown and environmental pollution: An Indian multi-state investigation. *Environmental Monitoring and Assessment*, *194*(2), 1–15.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113.
- Meinshausen, M., Lewis, J., McGlade, C., Gütschow, J., Nicholls, Z., Burdon, R., & Hackmann, B. (2022). Realization of Paris Agreement pledges may limit warming just below 2° C. *Nature*, *604*(7905), 304–309.
- Mishra, N., & Singh, A. (2018). Use of twitter data for waste minimisation in beef supply chain. *Annals of Operations Research*, *270*(1), 337–359.



- Mostafa, M. K., Gamal, G., & Wafiq, A. (2021). The impact of COVID 19 on air pollution levels and other environmental indicators-A case study of Egypt. *Journal of Environmental Management*, 277, 111496.
- Naeem, M. A., Karim, S., Yarovaya, L., & Lucey, B. M. (2023). COVID-induced sentiment and the intraday volatility spillovers between energy and other ETFs. *Energy Economics*, 122, 106677.
- Niu, B., Shen, Z., & Xie, F. (2021). The value of blockchain and agricultural supply chain parties' participation confronting random bacteria pollution. *Journal of Cleaner Production*, 319, 128579.
- Rathore, A. K., Kar, A. K., & Ilavarasan, P. V. (2017). Social media analytics: Literature review and directions for future research. *Decision Analysis*, 14(4), 229–249.
- Razzaq, A., Sharif, A., Aziz, N., Irfan, M., & Jermisittiparsert, K. (2020). Asymmetric link between environmental pollution and COVID-19 in the top ten affected states of US: A novel estimations from quantile-on-quantile approach. *Environmental Research*, 191, 110189.
- Schmidt, C. G., Wuttke, D. A., Ball, G. P., & Heese, H. S. (2020). Does social media elevate supply chain importance? An empirical examination of supply chain glitches, Twitter reactions, and stock market returns. *Journal of Operations Management*, 66(6), 646–669.
- Seuring, S., & Gold, S. (2012). Conducting content-analysis based literature reviews in supply chain management. *Supply Chain Management: An International Journal*, 17(5), 544–555. <https://doi.org/10.1108/13598541211258609>
- Sharma, A., Adhikary, A., & Borah, S. B. (2020). Covid-19' s impact on supply chain decisions: Strategic insights from NASDAQ 100 firms using Twitter data. *Journal of Business Research*, 117, 443–449.
- Sim, J., & Kim, B. (2021). Regulatory versus consumer pressure and retailer responsibility for upstream pollution in a supply chain. *Omega*, 101, 102250.
- Tinguely, P., Shrestha, Y. R., & Von Krogh, G. (2020). How does your labor force react to COVID-19? Employing social media analytics for preemptive decision making. *California Management Review*.
- Vallet-Bellmunt, T., Martínez-Fernández, M. T., & Capó-Vicedo, J. (2011). Supply chain management: A multidisciplinary content analysis of vertical relations between companies, 1997–2006. *Industrial Marketing Management*, 40(8), 1347–1367. <https://doi.org/10.1016/j.indmarman.2011.03.002>
- Wang, G., Ding, P., Chen, H., & Mu, J. (2020). Green fresh product cost sharing contracts considering freshness-keeping effort. *Soft Computing*, 24(4), 2671–2691.
- Wang, X., Sethi, S. P., & Chang, S. (2022). Pollution abatement using cap-and-trade in a dynamic supply chain and its coordination. *Transportation Research Part e: Logistics and Transportation Review*, 158, 102592.
- Weiss, S. M., Indurkha, N., Zhang, T., & Damerou, F. (2010). *Text mining: predictive methods for analyzing unstructured information*. Springer Science & Business Media.
- Wu, X., Tian, Z., & Guo, J. (2022). A review of the theoretical research and practical progress of carbon neutrality. *Sustainable Operations and Computers*, 3, 54–66.
- Xie, Y., Zhao, Y., Chen, Y., & Allen, C. (2022). Green construction supply chain management: Integrating governmental intervention and public–private partnerships through ecological modernisation. *Journal of Cleaner Production*, 331, 129986.
- Yang, M., Chen, L., Msigwa, G., Tang, K. H. D., & Yap, P. S. (2021). Implications of COVID-19 on global environmental pollution and carbon emissions with strategies for sustainability in the COVID-19 era. *Science of the Total Environment*, 809, 151657.
- Zhang, A., Alvi, M. F., Gong, Y., & Wang, J. X. (2022). Overcoming barriers to supply chain decarbonization: Case studies of first movers. *Resources, Conservation and Recycling*, 186, 106536.
- Zhang, D., Zhou, L., & Lim, J. (2020). From networking to mitigation: The role of social media and analytics in combating the COVID-19 pandemic. *Information Systems Management*, 37(4), 318–326.