Examining Public Perceptions of UK Rail Strikes: A Text Analytics Approach Using Twitter Data

Kyra Dong[*], Ying Kei Tse

Cardiff Business School, Cardiff University, CF10 3EU Cardiff, United Kingdom

[*] Correspondence: Kyra Dong (DongS14@cardiff.ac.uk)

Abstract: Social media, particularly Twitter, has emerged as a vital platform for understanding public opinion on contemporary issues. This study investigates public attitudes towards UK rail strikes by analyzing Twitter data and provides a framework to assist policymakers in the RMT Union and the government in managing social media information. A dataset comprising tweets related to rail strikes from 25 June 2022 to 7 October 2022 was collected and multidimensional scaling and sentiment analysis techniques were employed to examine public opinions and sentiments. The analysis revealed that the predominant trends in tweets were dissatisfaction and negativity, with users expressing inconvenience caused by the rail strikes. Interestingly, the public also questioned the government's capabilities, with some suggesting that rail strikes were politically motivated events orchestrated by the government. Sentiment analysis results indicated that approximately 85% of tweets displayed negative sentiment towards the rail strikes. This research contributes to the understanding of public attitudes derived from tweet mining and offers valuable insights for academics and policymakers in interpreting public reactions to current events. Based on the findings, recommendations for the RMT Union are proposed through the lenses of stakeholder orientation theory and signaling theory. For instance, fostering public engagement can help reduce information asymmetry between the RMT Union and the public, enabling the union to better comprehend public sentiment towards rail strikes. The approach amalgamates these two theories, presenting a novel theoretical perspective for such investigations and extending their applicability, while also providing clear and in-depth recommendations for the RMT Union.

Keywords: Social media, Text analytics, Rail strikes, Public opinion, Twitter mining, Information asymmetry, Stakeholder orientation theory

1 Introduction

Social networks, as a prominent type of Internet-based media, have emerged as an increasingly important source of information for many individuals. Social media platforms such as Facebook and Twitter grant users instant and open access to news and narratives, allowing them to build networks, interact, and share opinions. Recent studies have revealed that people are increasingly using social media applications for various purposes, including making new friends, socializing with old acquaintances, receiving information, and expressing opinions on current events [1, 2]. Opinions disseminated through social networks play a significant role in influencing public behaviors and attitudes in various aspects, such as purchasing products, predicting stock market trends, voting for presidents, and evaluating news [3, 4]. Consequently, academic attention has been drawn to the utilization of social media data in areas like customer relationship management, opinion tracking, and text filtering [5–7]. However, relatively few social media studies have concentrated on public attitudes and opinions regarding current events, particularly those concerning public transport strikes, which are highly sensitive issues.

In the United Kingdom, the National Union of Rail, Maritime and Transport Workers (RMT Union) announced a rail strike following unsuccessful talks with the government on June 20, 2022. The strike, expected to involve ten of thousands of workers and cause disruptions or delays on nearly every major line in England, Scotland, and Wales, commenced on June 21 and is predicted to persist for six months or longer if no agreement is reached. It is crucial for RMT Union to understand public opinion on this event to timely adjust the direction of activities, reach a consensus with public opinion, and achieve its ultimate goal. Therefore, analyzing public attitudes towards current affairs using

social media data and controlling public opinion in a timely manner are essential first steps in determining whether the RMT Union's rail strike can succeed.

This study aims to fill gaps in the literature and contribute to RMT Union's understanding of public opinion regarding the rail strike by investigating public perceptions and attitudes towards strike-related news on Twitter. Twitter data has been employed in various fields in past research, mainly for purposes like sentiment analysis [3, 8–10], market intelligence gathering, topic detection, and stock market insights [11–18]. The number of Twitter users in the UK is estimated at 15.8 million as of 2021 [19], and given its popularity, Twitter generates a vast amount of legally available data [20, 21]. This data includes user records, tweets, and metadata that facilitate tracking different types of tweet activity. Additionally, Twitter's unique short text format requires users to express their thoughts clearly and concisely [22–24]. Consequently, Twitter provides a means to explore user opinions [25, 26].

Text mining, which encompasses techniques such as data mining, machine learning, and information retrieval, is often employed to overcome the challenge of analyzing overloaded unstructured text [27–29]. Sentiment analysis, for example, enables automatic identification of sentiment in a large number of tweets, classifying them as positive, neutral, or negative [30, 31]. This study collected Twitter data and applied QDA Miner software to analyze public attitudes towards the rail strike, offering practical significance for RMT Union to control the event's direction and possibly influencing the success of the rail strike. Therefore, the following research questions are posed:

RQ1: What is the public's attitude toward the rail strike?

RQ2: What are the main concerns and interests of the public about the rail strike?

RQ3: How does text mining provide public opinion insights for RMT Union to better control the course of action?

The remainder of this study is structured as follows: Section 2 reviews the literature related to social media, customer and public engagement with social media, and text-mining research; Section 3 presents the research methods, combining sentiment analysis and clustering techniques; Section 4 provides data analysis and discusses the research results; Section 5 concludes the study, explaining its practical significance and limitations.

2 Literature Review

2.1 Social Media Research

The increasing utilization of social media by the public has led to its growing application in the business environment. Social media tools, as defined by Malita [26], facilitate the socialization of content and encourage collaboration, interaction, and communication through discussion, feedback, voting, comments, and sharing of information from all interested parties. Typically, social media refers to specific platforms through which people communicate, such as blogs, social networks, and multimedia websites, with popular examples including Facebook, Twitter, and YouTube [27, 28]. Due to the advantages of timely response and direct communication, social media has been extensively studied by practitioners and researchers. Investigations have focused on people's comments and interactions on specific platforms to address practical problems, especially the attitudes and communication of the public and consumers on social networking sites [29, 30]. Oh et al. [16] emphasized the importance of social media in driving customer engagement, as it transforms online users from passive consumers into active content generators who share information with others. Chevalier and Mayzlin [31] examined customer feedback on relative book sales in online stores and found that negative reviews (i.e., one-star reviews) have a greater impact than positive reviews (i.e., five-star reviews). They concluded that superior book reviews could lead to higher relative sales at the websites. Furthermore, Ayyub et al. and Snead [32, 33] highlighted that social media, based on the Internet, enables users to create, share, and exchange information. Since the public's attitude toward events or things influences their behavior, understanding their views and emotional tendencies can facilitate a better exploration of public behavior. Thus, customer and public engagement have been given new impetus with the development of social media platforms. However, it should be noted that social media data, which can reflect sentiment not captured by traditional data collection methods (e.g., questionnaires), is mostly limited to English. Moreover, most social media data reflects results under specific screening conditions, potentially limiting representativeness.

2.1.1 Customer and public engagement with social media

The concept of customer engagement is rooted in service-dominant logic [34]. Customer engagement is defined as "manifestations from customers toward a firm or a brand beyond purchase, resulting from motivational drivers" [16]. Many companies have shifted promotional resources from traditional media to digital platforms, interacting directly with customers [35, 36]. Goh et al. [37] determined that engaged customers' messages were 22 times more valuable than those of marketers, underscoring the importance of understanding customer engagement. Social media-driven customer engagement refers to a user's performance (beyond purchase) toward a company or brand in a social media environment, driven by social media availability and satisfaction [38–43]. Public engagement, broadly defined as the participation of citizens in public affairs [12], seeks to establish a relationship between local governments and citizens that extends beyond simple information exchange. The goal of public engagement is to support public interaction

and participation, leading to more informed government decisions [44]. The rapid development of social media has rendered it a powerful tool for reinforcing public engagement. Specifically, users can become active content creators instead of passive information recipients, stimulating public contribution to public life [45]. Additionally, real-time communication allows local governments and the public to engage in dialogue on social media platforms, fostering public engagement [46–48]. Analyses of public engagement in social media campaigns on Facebook, Twitter, and blogs by Cho et al. and Paek et al. [49, 50] demonstrated that people's use of each social media platform significantly affects their interaction with the campaign, which in turn mediates the relationship between social media use and offline communication behavior patterns.

2.2 Text Mining

Text mining, a form of data mining, has been applied across various disciplines to identify hidden patterns in large amounts of text data, particularly data from social media sources, such as categorizing negative or positive comments from consumers [51–53], exploring and tracking public political preferences [54], and determining customer satisfaction [46]. Text mining focuses on discovering useful models, trends, patterns, or rules from unstructured textual data, such as text files, HTML files, chat logs, and emails [55, 56]. As an automated technology, text mining can be employed to "effectively and systematically identify, extract, manage, integrate, and utilize knowledge in texts" [57]. Unlike traditional content analysis, text mining is primarily data-driven, aiming to automatically identify hidden patterns or trends in data [58] and subsequently generate explanations or models to elucidate interesting patterns and trends in text mining [59]. Sentiment analysis, a form of text mining, is widely used to analyze textual data.

2.2.1 Sentiment analysis

Sentiment analysis is frequently employed to examine language features from a sentiment perspective. This technique assists researchers in detecting sentiment polarity from textual data, helping to elucidate the relationship between public opinion formation and events [32]. As public attitudes toward issues influence behavior, it is essential to understand people's opinions and emotional tendencies in a timely manner. The proliferation of social media use has provided researchers with new methods and online resources to further explore public or consumer behavior. Sentiment analysis has recently been used to extract opinions from consumer product reviews [60], categorize positive and negative consumer product reviews [61], track emotional trends in online discussion forums [9], and monitor political views [46, 62]. Kuang and Yu et al. [40, 63] employed sentiment analysis to study the distribution of public sentiment in news articles during the COVID-19 pandemic. Moreover, numerous researchers have contributed to news sentiment analysis using various methods. Boudiukh et al. [64] utilized sentiment analysis to identify essential information in public news. The analysis was conducted at the sentence level, and a dynamic dictionary with predefined positive and negative words was employed to aid in determining sentiment polarity.

2.2.2 Cluster analysis

Cluster analysis is a crucial tool for automatically organizing and searching for information (such as unexpected trends) from unstructured text data [65, 66]. Although clustering is a technique used to group similar documents, it is not based on predefined topic classifications but dynamic clustering groupings [57]. Ying and Feng [67] employed data mining methods to analyze customers in the securities industry from the perspectives of customer value and customer behavior. They asserted that the clustering algorithm could serve as a standard customer segmentation method in data mining. Zhang [68] and Zhao et al. [69] used cluster analysis to classify consumer behavior characteristics, aiming to achieve intelligent product recommendations. In conclusion, cluster analysis is beneficial for practitioners and researchers who wish to explore customer segmentation through non-traditional methods.

3 Methodology

3.1 Data Collection and Preprocessing

The data utilized in this study were primarily collected from tweets and comments made by the public in response to rail strike news on Twitter between 25/06/2022 and 7/10/2022. The focus on this period was due to the heightened activity of public opinions and attitudes on Twitter within a week following the train strike, providing a large, effective, and timely dataset. The ExportComments tool was employed to extract 6,738 tweets using the search format: "Rail Strike" lang: en until: [end date] since: [start date]. As tweets contain abundant information but are considered unstructured text, data preprocessing was required to transform the text into a format suitable for analysis by machine learning algorithms.

Initially, URL links in the tweets were removed, and all data were converted to text format. Tokenization was then performed, dividing the text into discrete words. Subsequently, stop words (e.g., is, and, if, etc.) were filtered out. The resulting dataset comprised 2,678 eligible tweets. To import the data into QDA Miner software, all lowercase words were converted to uppercase. Following the data preparation process, QDA Miner software was utilized to

generate the public's attitude towards the rail strike event in the tweets and to determine the word frequency in the opinions.

Ethical considerations were taken into account when handling social media data in this study. Although the data were collected from Twitter, an open access platform, they were treated confidentially and were accessible only to the authors of this study. Furthermore, the data will be deleted six months after the conclusion of the study to prevent any ethical issues.

3.2 Multidimensional Scaling

This study employed a multidimensional scaling method to comprehensively analyze the co-occurrence of keywords within the dataset. Co-occurrence was defined as the occurrence of two words within the same context. To quantify the degree of co-occurrence, Jaccard's coefficient was employed as a measure of similarity between word pairs (i and j) [69]. The formula for Jaccard's coefficient is shown as follows:

$$\text{sim}_{\text{Jaccard}}\left(i,j\right) = \frac{a}{a+b+c}$$

In this context, the variable a represents the number of cases in which both words i and j are present, b denotes the number of cases where only word i is present, and c indicates the number of cases where only word j is present. Jaccard's coefficient is utilized as a measure of similarity, assigning equal weight to both matches and non-matches, while assigning zero weight to instances where neither word occurs (0-0 matches). A Jaccard coefficient of 0 indicates no similarity between the two words, whereas a coefficient of 1 signifies an exact match between them.

The Jaccard coefficient was utilized to measure similarity, and concept maps (as depicted in Figure 1) were generated as visual representations of the calculated closeness values for keywords using multidimensional scaling. In the concept map, each node represents a keyword, with its size indicating its frequency of occurrence. The distances between keywords on the map reflect the likelihood of certain terms appearing together. Keywords positioned close to each other on the map indicate a higher probability of co-occurrence, whereas keywords that are independent or have infrequent co-occurrence are placed farther apart.

3.3 Sentiment Analysis

Text mining techniques were applied in this study to automatically identify and classify patterns from large datasets and generate insights from unstructured text corpora. Sentiment analysis, a method used to extract subjectivity and polarity from textual data, was employed to analyze the tweets [70]. The Lexicon-based approach, which annotates tweets using a specific dictionary of words to determine polarity, was selected for this study.

The WordStat sentiment dictionary from QDA Miner software was applied, which was designed by combining negative and positive words from the "Harvard IV" and the "Linguistic and Word Count" dictionaries [71]. This general-purpose dictionary, containing over 9,526 negative word patterns and 4,669 positive word patterns, has been used in various settings, such as corporate sustainability reports [72], academic articles [73], and policy documents [74]. Separate scores for positive and negative sentiment were provided, each calculated based on terms containing thousands of word patterns. The semantic orientation value of tweets was calculated by the overall polarity (polarity = positive counts - negative counts), resulting in the final extreme value. A positive score was marked as positive, a negative score as negative, and a text with a score of "0" as neutral.

However, the WordStat sentiment dictionary has limitations when applied to specific domains, as some studies have shown that nearly three-quarters of the negative words in the dictionary are not typically negative in a financial context [75]. For example, words such as "mine", "tier", or "capital" may lead to a high rate of misclassification. Therefore, when using this dictionary, it is essential to focus on specific industry sectors and adapt it accordingly by adding domain-specific emotive words or phrases [76].

4 Results

Data from the dataset was input into QDA Miner software for text mining. Following preprocessing of the data as described in Section 3, a word list was generated. Table 1 displays the frequency of each word, with words such as "STRIKE", "RAIL", "SUPPORT", and "LABOUR" appearing most frequently in rail strike news. Despite its simplicity, word count analysis can provide insights into identifying hot topics and predicting key features of topics [23]. A proximity plot was employed to examine relevant topics related to the research focus and to provide an overview of the keywords associated with rail strike tweets. Words like "WORKER" and "GOVERNMENT" also appeared with high frequency, indicating the broad scope of the rail strike and its significance as a news event that garnered considerable attention. Frequent words such as "PAY", "COST", and "BLAME" further reflect public reactions and concerns regarding the issue. Word frequency analysis may thus assist practitioners (e.g., RMT Union) in rapidly identifying main topics on social media to investigate public attitudes towards rail strikes. Keywords such
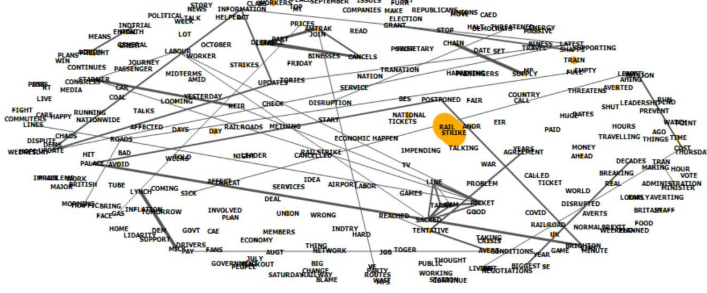
Table 1. Word frequency

	FREQUENCY	SHOWN	PROCESSED	TOTAL
STRIKE	2929	25.06%	8.38%	4.48%
RAIL	2846	24.25%	8.14%	4.35%
PEOPLE	286	2.45%	0.82%	0.44%
SUPPORT	250	2.14%	0.71%	0.38%
LABOUR	217	1.86%	0.62%	0.33%
WORKER	194	1.66%	0.55%	0.30%
TRAIN	178	1.52%	0.51%	0.27%
GOVERNMENT	171	1.46%	0.49%	0.26%
DAY	153	1.31%	0.44%	0.23%
TIME	132	1.13%	0.38%	0.20%
PAY	125	1.07%	0.36%	0.19%
GOOD	112	0.96%	0.32%	0.17%
RMT	102	0.87%	0.29%	0.16%
WORK	102	0.87%	0.29%	0.16%
HOME	98	0.84%	0.28%	0.15%
UK	95	0.81%	0.27%	0.15%
PARTY	89	0.76%	0.25%	0.14%
NATIONAL	86	0.74%	0.25%	0.13%
BREXIT	83	0.71%	0.24%	0.13%
UNION	77	0.66%	0.22%	0.12%
PUBLIC	75	0.64%	0.21%	0.11%
ANOR	74	0.63%	0.21%	0.11%
STAFF	74	0.63%	0.21%	0.11%
COUNTRY	66	0.56%	0.19%	0.10%
COST	64	0.55%	0.18%	0.10%
TOMORROW	61	0.52%	0.17%	0.09%
TRAVEL	58	0.50%	0.17%	0.09%
STOP	57	0.49%	0.16%	0.09%
WEEK	56	0.48%	0.16%	0.09%
NETWORK	55	0.47%	0.16%	0.08%
LINE	54	0.46%	0.15%	0.08%
MORNING	53	0.45%	0.15%	0.08%
YEAR	53	0.45%	0.15%	0.08%
PM	49	0.42%	0.14%	0.07%
YESTERDAY	48	0.41%	0.14%	0.07%
ACTION	47	0.40%	0.13%	0.07%
AFFECTED	47	0.40%	0.13%	0.07%
LIVING	46	0.39%	0.13%	0.07%
SERVICE	46	0.39%	0.13%	0.07%
BLAME	45	0.39%	0.13%	0.07%
CALLED	45	0.39%	0.13%	0.07%
CAR	45	0.39%	0.13%	0.07%
PICKET	45	0.39%	0.13%	0.07%
WRONG	45	0.39%	0.13%	0.07%
AIRPORT	44	0.38%	0.13%	0.07%
GREAT	44	0.38%	0.13%	0.07%
EMPTY	43	0.37%	0.12%	0.07%
JOB	43	0.37%	0.12%	0.07%
LEADER	42	0.36%	0.12%	0.06%
POLITICAL	42	0.36%	0.12%	0.06%
POOR	40	0.34%	0.12%	0.06%

as "STRIKE", "RAIL", "LABOUR", and "SUPPORT" predominantly appeared in tweets related to rail strikes, with "RMT", "STOP", and "MEDIA" also being identified as high-frequency words associated with rail strikes in tweets. Multidimensional scaling analysis was conducted, considering different units of investigation such as the co-

occurrence of keywords in paragraphs, sentences, documents, and archives. As this research focus was on tweets, the investigation unit was defined as a tweet. Concept maps were generated as graphic representations of proximity values computed using multidimensional scaling for all included keywords. Jaccard's coefficient was employed to measure similarity in the concept maps. Figure 1 illustrates the concept maps, where each node represents a keyword, and the size of the node indicates the frequency of the keyword's usage [77]. In Figure 1, the distinct clusters of cases can be identified based on the proximity of circles, represented by their distance and color, as well as the strength of their relationships, indicated by the number of connecting lines. This allows for the visualization and capturing of influential tweets. For instance, the focal group containing keywords "rail" and "strike" appears to be closely related to the "Crisis group".


Figure 1. Concept map of keywords

In summary, the following clustered groups were determined in this study (see Figure 2 – Figure 8 below):

A. The focal group which has the highest word frequency and contains the greatest number of tweets (59.97%) — "rail" "strike" — tweet example: Wasn't the government telling people to work from home cos the rail strike? Hypocrisy.

B. Neutral group: "non-event" — tweet example: Rail Strike has been a total non-event in my part of the World.

C. Support group: "support" — tweet example: Massive support for rail strikes amongst both rail workers public.

D. Crisis group: "crisis" "pay rise"— tweet example: Rail strike. If there is a rail strike, we are going to see a supply chain crisis like nothing in our history.

E. Complain group: "annoying" "noise" — tweet example: Very annoying rail strike.

F. Video group: "Video" "slogan" — tweet example: Govt's argument is a red herring all from empty slogans and no solutions, against his articulate precise framing of the situation (a video of Mick Lynch).

G. Labour party group: "vote" "labour party" — tweet example: Vote for Labour Party who support Rail strikes which will put up prices, and cost jobs businesses.

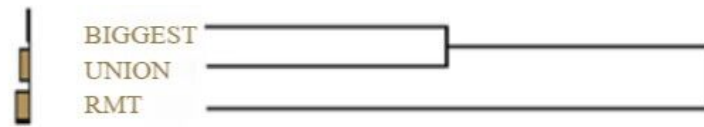**Figure 2.** Focal group (average sentiment score: -0.357)
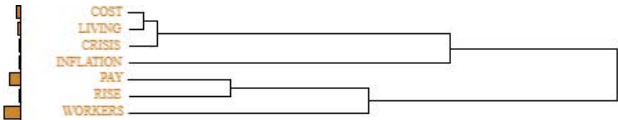
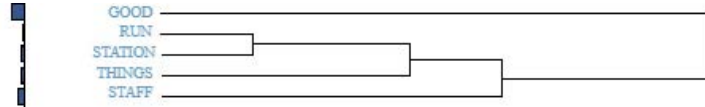**Figure 3.** Neutral group (average sentiment score: -0.034)
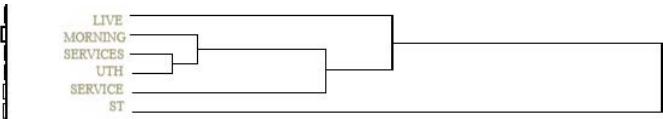
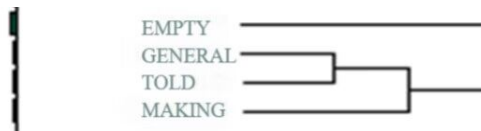**Figure 4.** Support group (average sentiment score: 0.595)

Clustered analysis was chosen for this study as it allows for more precise grouping of words, rather than finding synonyms for keywords or substitutes for single words. For example, the "Video group" comprises all topics and
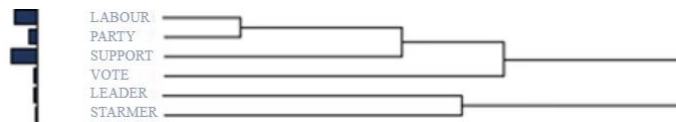
Figure 5. Crisis group (average sentiment score: -0.121)

Figure 6. Complain group (average sentiment score: -0.163)

Figure 7. Video group (average sentiment score: -0.127)

Figure 8. Labour party group (average sentiment score: -0.232)

keywords related to video, enabling more accurate exploration of the relationships among keywords within the group. In addition, analysis of the seven clustered groups revealed that all but the "Neutral group" reflected part of public opinions on the rail strike on Twitter. Through in-depth analysis of public tweets, it was found that some members of the public not only paid close attention to videos related to the rail strike but also expressed dissatisfaction with the government for not providing practical solutions or offering only "empty slogans" and "no solutions". Furthermore, some individuals questioned the government's problem-solving abilities and viewed the rail strike as a politically driven event.

For the lexicon-based sentiment analysis, the WordStat sentiment dictionary [77] was employed, which calculates semantic orientation values by subtracting the number of negative words from the number of positive words in a tweet. A positive sentiment was assigned to tweets with a semantic orientation value greater than 0, while a negative sentiment was assigned to those with a value less than 0. Tweets with a value equal to 0 were considered neutral. Among the 2,687 tweets analyzed in this study, 2,228 were negative, 122 were positive, and 337 were neutral. As suggested by Li et al. [78], focus can be narrowed to only the number of positive and negative tweets when the proportion of neutral tweets in the dataset is lower than the number of positive tweets. Figure 9 illustrates the sentiment analysis results, showing that negative sentiment was far more prevalent than positive sentiment in tweets about the rail strike.

"Tweet 1: There is a rail strike people like to want to make our lives a misery. -1"

"Tweet 2: Do not support workers in rail strike? A labour leader who does not back workers is no labour leader. -1"

"Tweet 3: Is not it about time passengers had a rail strike against the most outrageous ticket fares. . . it needs to happen! -3"

According to the WordStat sentiment dictionary, in "Tweet 1", there are two negative counts (2) and one positive count (1). Therefore, the sentiment orientation value = positive count (1) - negative counts (2) = -1, which indicates that the sentiment of "Tweet 1" is negative. Similarly, in "Tweet 2", the negative count is 1 and the positive count is 0. Thus, the sentiment orientation value is -1, meaning the sentiment of "Tweet 2" is negative. In "Tweet 3", there are three negative counts and no positive counts. Consequently, "Tweet 3" has a sentiment orientation value of -3, signifying that the sentiment of "Tweet 3" is negative. The results suggest that, in the context of the rail strike, the public's attitude is predominantly negative. We will present the average sentiment score for each cluster group in Figure 2-Figure 8.

The analysis of the results revealed that nearly 85% of the tweets in the dataset expressed negative sentiment towards the rail strike. This finding is consistent with previous research indicating that social media users tend

to express negative emotions when discussing controversial events or crises [79, 80]. The negative sentiment was particularly evident in the "Crisis group" and "Complain group". However, it is important to note that the "Support group" exhibited a higher average positive sentiment score, suggesting that there is still a portion of the public who supports the rail workers or their demands.

Distribution of keywords (Frequency)
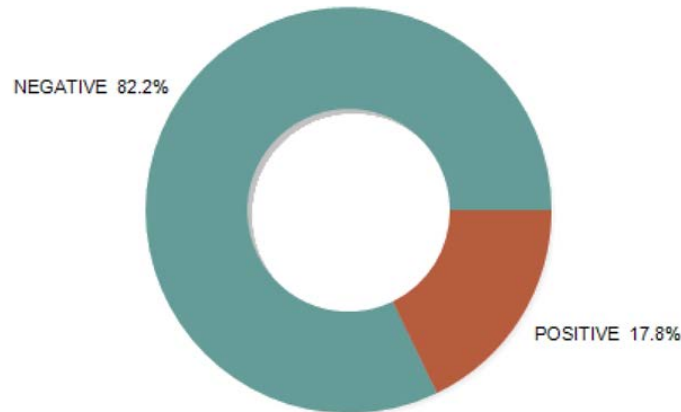
Figure 9. Result of sentiment analysis

5 Discussion

In analyzing the rail strike event, the survey results reveal a predominantly negative public attitude towards the RMT Union action. The number of negative tweets (2228) significantly outweighs the positive ones (459), indicating minimal public support for the strike. A multidimensional scaling method was employed to identify seven distinct groups in the sample, namely "Focal", "Neutral", "Support", "Crisis", "Complain", "Video", and "Labour party". These clusters were formed due to the high co-occurrence of elements within each group. The majority of these groups expressed their negative views on the rail strike directly via Twitter. Notably, the "Focal group" contained the largest number of tweets (59.97%) and frequently used terms such as "hypocrisy" to convey their negative stance.

5.1 Application of Signalling Theory and Stakeholder Orientation Theory

The application of signalling theory and stakeholder orientation theory was utilized to provide more insightful suggestions to the RMT Union and address the research question (RQ3). According to Karaman et al. [79], signalling theory encompasses three subjects: sender, receiver, and signal. Within the context of this study, the sender is the RMT Union, the signal is news about the rail strike on Twitter, and the public serves as the receiver, providing feedback through tweets. The focus of signalling theory is on information asymmetries among multiple entities, such as individuals or organizations. Consequently, public engagement is more likely to reduce information asymmetry between the RMT Union and the public, assisting the RMT Union in understanding the public's stance on rail strikes and garnering greater support.

Stakeholder orientation, as described by Brulhart et al. [81], is a valuable, non-imitable, and irreplaceable resource that can confer a competitive advantage. It refers to an organization's capacity to establish and maintain long-term relationships with various stakeholder categories. In cyberspace, public groups can discuss and influence each other's perceptions and opinions on a particular issue, which can be perceived as a resource or knowledge for the organization. Consequently, organizations should regard the public as stakeholders who can readily share their ideas and opinions about the company and its activities on social media platforms.

Focusing on the study results, the "Neutral group" exhibited an average sentiment score of -0.034. As they remained neutral about the signals, it can be inferred that the RMT Union's signal did not resonate with them. The "Support group," although seemingly the most important group identified through data mining, may be the last group to which the RMT Union should allocate resources for maintaining the signalling process. The average sentiment score for this group was 0.595, and most Twitter users within it were RMT Union members who openly supported the rail strike. By examining the content of tweets from this stakeholder group, the RMT Union could potentially extract valuable information and knowledge to shape its industrial action strategy based on the opinions of key stakeholders on social media.

The "Focal group" contained the highest word frequency and had the largest number of tweets (59.97%). With an average sentiment score of -0.357, this group demonstrated a negative attitude towards rail strikes. This finding suggests that the "Focal group" was highly responsive to the RMT Union's signal. Additionally, the average sentiment score of the "Complain", "Crisis", and "Labour party" groups were -0.163, -0.121, and -0.232, respectively. Despite their complaints and predictions that rail strikes would lead to crises and increased train fares, this information cycle contributed to a heightened sense of public engagement, attracting more attention and retweets. As a consequence, the number of signals increased, and these stakeholders generated resources and knowledge regarding the industrial actions.

According to Donaldson and Preston [82], all stakeholders merit attention, regardless of their ability to create wealth or the value of the resources they provide. This implies that even if stakeholders bring negative or positive value to the organization, their opinions and feedback should be acknowledged. Organizational managers should also consider the influence of stakeholders on broader societal strata. Although five of the seven groups displayed negative attitudes towards rail strikes, monitoring and managing public sentiment can contribute to the success of the RMT Union's actions.

5.2 Suggestions for RMT Union

Social media platforms facilitate stakeholder engagement by offering opportunities for stakeholder groups to stay informed, identify common interests, express and share opinions and demands, and organize and coordinate interventions. Stakeholder engagement, characterized by the interaction between a company and its stakeholders, plays a crucial role in helping organizations obtain information from them. Social media platforms are particularly effective for stakeholder engagement, as they enable organizations to communicate with stakeholders and receive feedback through various responses (e.g., likes, shares, and comments).

In a stakeholder-oriented organization, key groups such as employees, customers, governments, unions, shareholders, and executives are explicitly considered when shaping and determining strategic actions. Moreover, an organization should serve multiple stakeholders to not only secure their support but also acquire resources, knowledge,and legitimacy. This approach can contribute to the long-term success of the organization. To address the RMT Union's current situation, the following suggestions are proposed:

(1) Develop a comprehensive social media strategy: The RMT Union should establish a social media strategy that encompasses monitoring, engagement, and communication with stakeholders. This strategy should focus on understanding public sentiment, addressing concerns, and disseminating accurate information to counter misinformation and negative perceptions.

(2) Employ social media monitoring tools: Utilizing social media monitoring tools can help the RMT Union track public sentiment and engagement related to their activities. These tools can aid in identifying trends, influential users, and popular topics, enabling the Union to efficiently address stakeholders' concerns and needs.

(3) Create targeted content: The RMT Union should create and share content that specifically targets various stakeholder groups. This content may include educational materials that explain the reasons for strikes, the Union's goals, and the benefits of their actions in the long term. By addressing the concerns of different stakeholder groups, the Union can foster greater understanding and support.

(4) Engage in transparent dialogue: The RMT Union should actively participate in conversations with stakeholders on social media platforms. By engaging in transparent and open dialogue, the Union can build trust, address misconceptions, and clarify its stance on various issues.

(5) Leverage influential supporters: The RMT Union should identify influential supporters within their stakeholder groups and collaborate with them to amplify their message. These supporters can act as advocates for the Union's cause and help counterbalance negative sentiment.

(6) Establish partnerships with other organizations: The RMT Union should seek partnerships with other organizations that share similar objectives or can provide support in achieving their goals. Such partnerships can enhance the Union's credibility and create a united front in addressing stakeholder concerns.

(7) Continuously evaluate and adapt the strategy: The RMT Union should regularly assess the effectiveness of its social media strategy and adapt it based on stakeholder feedback, public sentiment, and emerging trends. This iterative process will help the Union stay responsive to stakeholder needs and remain agile in the face of changing circumstances.

By applying these suggestions, the RMT Union can strengthen its stakeholder engagement and communication efforts, leading to better understanding and support for its actions. Furthermore, these strategies can contribute to the long-term success of the organization and its ability to serve multiple stakeholders effectively.

6 Conclusions

In light of the rail strike's commencement, an analysis of 2687 tweets was conducted to investigate public opinions and attitudes towards the issue, as expressed on social media platforms. A hybrid text mining approach,

encompassing sentiment analysis and cluster analysis, was employed. The primary objective of this research was to elucidate public attitudes and opinions regarding rail strikes. A key finding was that word frequency analysis swiftly identified potential thematic features, such as "LABOUR", "CRISIS", and "PUBLIC".

Utilizing a multidimensional scaling approach, seven distinct groups were identified, and a novel research methodology for exploring potential public groups was proposed. Sentiment analysis results indicated a predominantly negative perception of rail strikes, with approximately 85% of the analyzed tweets expressing negative sentiments. Thus, the RMT Union was able to comprehend the reasons behind this negativity, including concerns about government competence and potential impacts on employment opportunities, and create an action plan to further manage the situation by leveraging text mining insights.

This study offers an in-depth understanding of public attitudes towards rail strikes through text mining, providing valuable insights and a reference basis for scholars interested in employing text mining methods to explore public opinions on current events and their developmental trends. Furthermore, the integration of signalling theory and stakeholder orientation theory enhances the interpretation of the study's results and offers a novel theoretical perspective for future research. Comprehensive and practical data findings and strategies are provided for the RMT Union, such as improving communication with various stakeholder groups and addressing public concerns promptly during rail strikes. Additionally, this study presents insights into government policymaking from a political perspective based on data analysis results.

Despite its contributions, this study presents some limitations that warrant further exploration. Firstly, the research focused solely on the opinions of social media users, while neglecting the views of those who do not frequently use these platforms. Future studies could employ interview methods to capture a more accurate representation of the population's opinions. Secondly, the cluster analysis used to identify differing tweet groups did not delve into the reasons behind the tweets' content. It is recommended that future research adopt the public group segmentation and identification method proposed in this study to uncover hidden information within the clusters. Lastly, the lexicon-based sentiment analysis approach employed in this study has certain shortcomings, such as the difficulty in accurately annotating specialized words in specific fields [24, 78]. It is suggested that future research should focus on improving existing dictionaries or utilizing more accurate labeling methods to overcome these limitations.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] C. McCarthy, "Manage challenges, embrace opportunities of social-media world," *Campus Leg. Advisor*, vol. 15, no. 10, pp. 1–5, 2015. https://doi.org/10.1002/cala.30116

[2] J. Phua, S. V. Jin, and J. J. Kim, "Uses and gratifications of social networking sites for bridging and bonding social capital: A comparison of facebook, twitter, instagram, and snapchat," *Comput. Hum. Behav.*, vol. 72, pp. 115–122, 2017. https://doi.org/10.1016/j.chb.2017.02.041

[3] R. Santhiran and K. D. Varathan, "Feature based ranking approaches in opinion mining: A systematic review," *Adv. Sci. Lett.*, vol. 24, no. 3, pp. 1881–1884, 2018. https://doi.org/10.1166/asl.2018.11181

[4] L. Zhang, H. Yuan, and R. Y. K. Lau, "Predicting and visualizing consumer sentiments in online social media," in *IEEE 13th International Conference on e-Business Engineering (ICEBE 2016)*. IEEE, Macau, China, 2016, pp. 1–8. https://doi.org/10.1109/ICEBE.2016.025

[5] C. Zhang, D. Zeng, J. Li, F. Y. Wang, and W. Zuo, "Sentiment analysis of chinese documents: From sentence to document level," *J. Am. Soc. Info. Sci. Tech.*, vol. 60, no. 12, pp. 2474–2487, 2009. https://doi.org/10.1002/asi.21206

[6] G. Governatori and R. Iannella, "A modelling and reasoning framework for social networks policies," *Enterprise Infor. Syst.*, vol. 5, no. 1, pp. 145–167, 2010. https://doi.org/10.1080/17517575.2010.513014

[7] Z. C. Steinert-Threlkeld, D. Mocanu, A. Vespignani, and J. Fowler, "Online social networks and offline protest," *EPJ Data Sci.*, vol. 4, no. 1, 2015. https://doi.org/10.1140/epjds/s13688-015-0056-y

[8] C. K. H. Lee and Y. K. Tse, "Improving peer-to-peer accommodation service based on text analytics," *Ind. Manag. Data Syst.*, vol. 121, no. 2, pp. 209–227, 2020. https://doi.org/10.1108/imds-02-2020-0105

[9] P. Aragón, V. Gómez, and A. Kaltenbrunner, "Detecting platform effects in online discussions," *Pol. Int.*, vol. 9, no. 4, pp. 420–443, 2017. https://doi.org/10.1002/poi3.158

[10] M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter brand sentiment analysis: A hybrid system using N-gram analysis and dynamic artificial neural network," *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6266–6282, 2013. https://doi.org/10.1016/j.eswa.2013.05.057

[11] U. R. Hodeghatta and S. Sahney, "Understanding Twitter as an e-WOM," *J. Syst. Info. Tech.*, vol. 18, no. 1, pp. 89–115, 2016. https://doi.org/10.1108/jsit-12-2014-0074

[12] G. Rowe and L. J. Frewer, "A typology of public engagement mechanisms," *Sci., Tech., Hum. Values*, vol. 30, no. 2, pp. 251–290, 2005. https://doi.org/10.1177/0162243904271724

[13] Y. M. Li and T. Y. Li, "Deriving market intelligence from microblogs," *Decision Support Syst.*, vol. 55, no. 1, pp. 206–217, 2013. https://doi.org/10.1016/j.dss.2013.01.023

[14] Y. Lu, F. Wang, and R. Maciejewski, "Business intelligence from social media: A study from the vast box office challenge," *IEEE Comput. Graphics Appl.*, vol. 34, no. 5, pp. 58–69, 2014. https://doi.org/10.1109/mcg.2014.61

[15] J. Cigarran, A. Castellanos, and A. Garcia-Serrano, "A step forward for topic detection in Twitter: An FCA-based approach," *Expert Syst. Appli.*, vol. 57, pp. 21–36, 2016. https://doi.org/10.1016/j.eswa.2016.03.011

[16] H. Oh, A. Animesh, and A. Pinsonneault, "Free Versus For-a-Fee: The impact of a paywall on the pattern and effectiveness of word-of-mouth via social media," *MIS Quart.*, vol. 40, no. 1, pp. 31–56, 2016. https://doi.org/10.25300/misq/2016/40.1.02

[17] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *J. Lang. Soc. Psychol.*, vol. 29, no. 1, pp. 24–54, 2009. https://doi.org/10.1177/0261927x09351676

[18] J. Bhattacharjya, A. Ellison, and S. Tripathi, "An exploration of logistics-related customer service provision on Twitter," *Int. J. Phys. Distribution Logistics Manag.*, vol. 46, no. 6/7, pp. 659–680, 2016. https://doi.org/10.1108/ijpdlm-01-2015-0007

[19] B. K. Chae, "Insights from hashtag #supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research," *Int. J. Production Econ.*, vol. 165, pp. 247–259, 2015. https://doi.org/10.1016/j.ijpe.2014.12.037

[20] L. Dang Xuan, S. Stieglitz, J. Wladarsch, and C. Neuberger, "An investigation of influentials and the role of sentiment in political communication on Twitter during election periods," *Info., Comm. Soc.*, vol. 16, no. 5, pp. 795–825, 2013. https://doi.org/10.1080/1369118x.2013.783608

[21] Y. Gorodnichenko, T. Pham, and O. Talavera, "Social media, sentiment and public opinions: Evidence from #Brexit and #uselection," *Euro. Econ. Rev.*, vol. 136, p. 103772, 2021. https://doi.org/10.1016/j.euroecorev.2021.103772

[22] I. Ileri and P. Karagoz, "Detecting user emotions in Twitter through collective classification," in *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2016)*. SciTePress, Porto, Portugal, 2016, pp. 205–212. https://doi.org/10.5220/0006037502050212

[23] M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4241–4251, 2013. https://doi.org/10.1016/j.eswa.2013.01.019

[24] B. Li, K. C. Chan, C. Ou, and S. Ruifeng, "Discovering public sentiment in social media for predicting stock movement of publicly listed companies," *Info. Syst.*, vol. 69, pp. 81–92, 2017. https://doi.org/10.1016/j.is.2016.10.001

[25] Y. Halberstam and B. Knight, "Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter," *J. Pub. Econ.*, vol. 143, pp. 73–88, 2016. https://doi.org/10.1016/j.jpubeco.2016.08.011

[26] L. Malita, "Social media time management tools and tips," *Procedia Compu. Sci.*, vol. 3, pp. 747–753, 2011. https://doi.org/10.1016/j.procs.2010.12.123

[27] R. Guesalaga, "The use of social media in sales: Individual and organizational antecedents, and the role of customer engagement in social media," *Ind. Mar. Manag.*, vol. 54, pp. 71–79, 2016. https://doi.org/10.1016/j.indmarman.2015.12.002

[28] D. M. Boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," *IEEE Engi. Manag. Rev.*, vol. 38, no. 3, pp. 16–31, 2010. https://doi.org/10.1109/emr.2010.5559139

[29] J. Hwang, A. Eves, and J. L. Stienmetz, "The impact of social media use on consumers' restaurant consumption experiences: A qualitative study," *Sustainability*, vol. 13, no. 12, p. 6581, 2021. https://doi.org/10.3390/su13126581

[30] A. M. Gamboa and H. M. Gonçalves, "Customer loyalty through social networks: Lessons from Zara on Facebook," *Bus. Horiz.*, vol. 57, no. 6, pp. 709–717, 2014. https://doi.org/10.1016/j.bushor.2014.07.003

[31] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," *J. Mar. Res.*,

vol. 43, no. 3, pp. 345–354, 2006. https://doi.org/10.1509/jmkr.43.3.345

[32] K. Ayyub, S. Iqbal, M. W. Nisar, S. G. Ahmad, and E. U. Munir, "Stance detection using diverse feature sets based on machine learning techniques," *J. Intell. Fuzzy Syst.*, vol. 40, no. 5, pp. 9721–9740, 2021. https://doi.org/10.3233/jifs-202269

[33] J. T. Snead, "Social media use in the U.S. executive branch," *Gov. Info. Quar.*, vol. 30, no. 1, pp. 56–63, 2013. https://doi.org/10.1016/j.giq.2012.09.001

[34] S. Streukens, A. van Riel, D. Novikova, and S. Leroi-Werelds, "Boosting customer engagement through gamification: A customer engagement marketing approach," pp. 35–54, 2019. https://doi.org/10.4337/978178 8114899.00008

[35] M. Paruthi and H. Kaur, "Scale development and validation for measuring online engagement," *J. Int. Commer.*, vol. 16, no. 2, pp. 127–147, 2017. https://doi.org/10.1080/15332861.2017.1299497

[36] R. P. Schumaker, A. T. Jarmoszko, and C. S. Labedz, "Predicting wins and spread in the premier league using a sentiment analysis of Twitter," *Decision Support Syst.*, vol. 88, pp. 76–84, 2016. https://doi.org/10.1016/j.dss.2016.05.010

[37] K. Y. Goh, C. S. Heng, and Z. Lin, "Social media brand community and consumer behavior: Quantifying the relative impact of user- and marketer-generated content," *Info. Syst. Res.*, vol. 24, no. 1, pp. 88–107, 2013. https://doi.org/10.1287/isre.1120.0469

[38] J. Braojos, J. Benitez, and J. Llorens, "How do social commerce-IT capabilities influence firm performance? Theory and empirical evidence," *Info. Manag.*, vol. 56, no. 2, pp. 155–171, 2019. https://doi.org/10.1016/j.im.2018.04.006

[39] S. K. Kim, M. J. Park, and J. J. Rho, "Effect of the government's use of social media on the reliability of the government: Focus on Twitter," *Pub. Manag. Rev.*, vol. 17, no. 3, pp. 328–355, 2013. https://doi.org/10.1080/14719037.2013.822530

[40] W. Kuang, "Empirical studies on new-media public opinion," in *Social Media in China.* Springer, 2018, pp. 257–261. https://doi.org/10.1007/978-981-13-0914-4_10

[41] J. M. Andzulis, N. G. Panagopoulos, and A. Rapp, "A review of social media and implications for the sales process," *J. Pers. Sell. Sales Manag.*, vol. 32, no. 3, pp. 305–316, 2012. https://doi.org/10.2753/pss0885-3134 320302

[42] R. N. Bolton, "Customer engagement conceptualization and conceptual relationships," in *Handbook of Research on Customer Engagement.* Edward Elgar Publishing, 2019, pp. 114–125. https://doi.org/10.4337/978178881 14899.00012

[43] H. Y. Wong and B. Merrilees, "An empirical study of the antecedents and consequences of brand engagement," *Mkt. Intell. Plan.*, vol. 33, no. 4, pp. 575–591, 2015. https://doi.org/10.1108/MIP-08-2014-0145

[44] G. Rowe and L. J. Frewer, "Public participation methods: A framework for evaluation," *Sci., Tech., Hum. Values*, vol. 25, no. 1, pp. 3–29, 2000. https://doi.org/10.1177/016224390002500101

[45] S. S. Lee, D. Boshnakova, and J. Goldblatt, "Marketing with wikis, websites, blogs, and podcasts," in *The 21st Century Meeting and Event Technologies.* CRC Press, 2017, pp. 183–205. https://doi.org/10.1201/b19849-12

[46] D. Kang and Y. Park, "Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and vikor approach," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1041–1050, 2014. https://doi.org/10.1016/j.eswa.2013.07.101

[47] C. Caba Pérez, M. P. Rodríguez Bolívar, and A. M. López Hernández, "The use of web 2.0 to transform public services delivery: The case of Spain," *Pub. Adm. Info. Tech.*, pp. 41–61, 2012. https://doi.org/10.1007/978-1-4614-1448-3_4

[48] J. F. F. E. Araujo and F. Tejedo-Romero, "Local government transparency index: Determinants of municipalities' rankings," *Int. J. Pub. Sect. Manag.*, vol. 29, no. 4, pp. 327–347, 2016. https://doi.org/10.1108/IJPSM-11-2015-0199

[49] M. Cho, T. Schweickart, and A. Haase, "Public engagement with nonprofit organizations on Facebook," *Pub. Relat. Rev.*, vol. 40, no. 3, pp. 565–567, 2014. https://doi.org/10.1016/j.pubrev.2014.01.008

[50] H. J. Paek, T. Hove, Y. Jung, and R. T. Cole, "Engagement across three social media platforms: An exploratory study of a cause-related PR campaign," *Pub. Relat. Rev.*, vol. 39, no. 5, pp. 526–533, 2013. https://doi.org/10.1016/j.pubrev.2013.09.013

[51] S. C. Ketron, J. A. Siguaw, and X. Sheng, "Maladaptive consumer behaviors and marketing responses in a pandemic," *ICT Evol. Work*, pp. 27–48, 2021. https://doi.org/10.1007/978-981-33-4134-0_2

[52] C. S. B. Ngai, R. G. Singh, W. Lu, and A. C. Koon, "Grappling with the COVID-19 health crisis: Content analysis of communication strategies and their effects on public engagement on social media," *J. Med. Int. Res.*, vol. 22, no. 8, 2020. https://doi.org/10.2196/21360

[53] A. Humphreys and R. J. H. Wang, "Automated text analysis for consumer research," *J. Consum. Res.*, vol. 44, no. 6, pp. 1274–1306, 2017. https://doi.org/10.1093/jcr/ucx104

[54] A. Ceron, L. Curini, S. M. Iacus, and G. Porro, "Every tweet counts? how sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France," *New Media Soc.*, vol. 16, no. 2, pp. 340–358, 2013. https://doi.org/10.1177/1461444813480466

[55] S. M. Shah, M. Lütjen, and M. Freitag, "Text mining for supply chain risk management in the apparel industry," *Appl. Sci.*, vol. 11, no. 5, p. 2323, 2021. https://doi.org/10.3390/app11052323

[56] C. L. Yang and T. P. Q. Nguyen, "Constrained clustering method for class-based storage location assignment in warehouse," *Ind. Manag. Data Syst.*, vol. 116, no. 4, pp. 667–689, 2016. https://doi.org/10.1108/imds-09-2015-0361

[57] M. Lamba and M. Madhusudhan, "Tools and techniques for text mining and visualization," in *Text Mining for Information Professionals*. Springer, 2022, pp. 295–318. https://doi.org/10.1007/978-3-030-85085-2_10

[58] M. Galchenko, A. Gushchinsky, W. Izdebski, and J. Skudlarski, "Data mining and statistics methods for advanced training course quality measurement: Case study," *Found. Manag.*, vol. 6, no. 3, pp. 47–56, 2014. https://doi.org/10.1515/fman-2015-0017

[59] J. Guo, L. D. Xu, G. Xiao, and Z. Gong, "Improving multilingual semantic interoperation in cross-organizational enterprise systems through concept disambiguation," *IEEE Trans. Ind. Inform.*, vol. 8, no. 3, pp. 647–658, 2012. https://doi.org/10.1109/TII.2012.2188899

[60] S. Negi and P. Buitelaar, "Towards the extraction of customer-to-customer suggestions from reviews," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*. Association for Computational Linguistics, Lisbon, Portugal, 2015, pp. 2159–2167.

[61] V. D. Kaur, "Sentimental analysis of book reviews using unsupervised semantic orientation and supervised machine learning approaches," in *2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT)*. IEEE, Bangalore, India, 2018, pp. 1444–1448. https://doi.org/10.1109/ICGCIoT.2018.8753089

[62] F. Namugera, R. Wesonga, and P. Jehopio, "Text mining and determinants of sentiments: Twitter social media usage by traditional media houses in Uganda," *Comput. Soc. Netw.*, vol. 6, no. 1, 2019. https://doi.org/10.1186/s40649-019-0063-4

[63] X. Yu, C. Zhong, D. Li, and W. Xu, "Sentiment analysis for news and social media in COVID-19," *J. Comput. Soc. Sci.*, vol. 6, pp. 19–57, 2023. https://doi.org/10.1145/3423333.3431794

[64] J. Boudoukh, R. Feldman, S. Kogan, and M. Richardson, "Information, trading, and volatility: Evidence from firm-specific news," *Rev. Financ. Stud.*, vol. 32, no. 3, pp. 992–1033, 2019. https://doi.org/10.1093/rfs/hhy083

[65] M. U. Islam, F. B. Ashraf, A. I. Abir, and M. A. Mottalib, "Polarity detection of online news articles based on sentence structure and dynamic dictionary," in *2017 20th International Conference of Computer and Information Technology (ICCIT)*. IEEE, Dhaka, Bangladesh, 2017, pp. 1–6. https://doi.org/10.1109/ICCITECHN.2017.8281777

[66] C. C. Aggarwal, "Cluster analysis: Advanced concepts," in *Data Mining: The Textbook*. Springer, 2015, pp. 205–236. https://doi.org/10.1007/978-3-319-14142-8_7

[67] Y. Li and F. Lin, "Customer segmentation analysis based on SOM clustering," in *2008 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*. IEEE, Beijing, China, 2008, pp. 1926–1931. https://doi.org/10.1109/SOLI.2008.4686353

[68] Z. Zhang, "Cluster analysis of consumer's behaviors based on unsupervised learning," in *2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM)*. ACM, New York, USA, 2021, pp. 1178–1184. https://doi.org/10.1145/3495018.3495360

[69] C. Zhao, S. Wang, and D. Li, "Fuzzy sentiment membership determining for sentiment classification," in *2014 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE, Shenzhen, China, 2014, pp. 599–606. https://doi.org/10.1109/ICDMW.2014.137

[70] L. Wang, R. Xia, and H. Li, "Sentiment lexicon construction with representation learning based on hierarchical sentiment supervision," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. ACL, Copenhagen, Denmark, 2017, pp. 502–510. https://doi.org/10.18653/v1/D17-1052

[71] A. Mehto and K. Indras, "Data mining through sentiment analysis: Lexicon based sentiment analysis model using aspect catalogue," in *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*. Indore, India, 2016, pp. 1–6. https://doi.org/10.1109/CDAN.2016.7570901

[72] I. Lock and P. Seele, "The credibility of CSR (Corporate Social Responsibility) reports in Europe. evidence from a quantitative content analysis in 11 countries," *J. Cleaner Prod.*, vol. 122, pp. 186–200, 2016. https://doi.org/10.1016/j.jclepro.2016.02.060

[73] L. Mora, X. Wu, and A. Panori, "Mind the gap: Developments in autonomous driving research and the

sustainability challenge," *J. Cleaner Prod.*, vol. 275, p. 124087, 2020. https://doi.org/10.1016/j.jclepro.2020.1 24087

[74] J. Van Alstine and R. Barkemeyer, "Business and development: Changing discourses in the extractive industries," *Resour. Pol.*, vol. 40, pp. 4–16, 2014. https://doi.org/10.1016/j.resourpol.2014.01.006

[75] B. Thapa, "Sentiment analysis of cyber security content on Twitter and Reddit," *Data Min. Mach. Learn.*, 2020. https://doi.org/10.5121/csit.2022.120708

[76] "Sentiment dictionaries – what is sentiment analysis?" 2018. https://provalisresearch.com/products/content-an alysis-software/wordstat-dictionary/sentiment-dictionaries/

[77] Y. K. Tse, M. Zhang, B. Doherty, P. Chappell, and P. Garnett, "Insight from the Horsemeat Scandal: Exploring the consumers' opinion of tweets toward Tesco," *Ind. Manag. Data Syst.*, vol. 116, no. 6, pp. 1178–1200, 2016. https://doi.org/10.1108/IMDS-10-2015-0417

[78] X. Li, M. Xu, W. Zeng, Y. K. Tse, and H. K. Chan, "Exploring customer concerns on service quality under the COVID-19 crisis: A social media analytics study from the retail industry," *J. Retail. Consum. Serv.*, vol. 70, p. 103157, 2023. https://doi.org/10.1016/j.jretconser.2022.103157

[79] A. S. Karaman, M. Kilic, and A. Uyar, "Green logistics performance and sustainability reporting practices of the logistics sector: The moderating effect of corporate governance," *J. Cleaner Prod.*, vol. 258, p. 120718, 2020. https://doi.org/10.1016/j.jclepro.2020.120718

[80] S. A. Taj, "Application of signaling theory in management research: Addressing major gaps in theory," *Euro. Manag. J.*, vol. 34, no. 4, pp. 338–348, 2016. https://doi.org/10.1016/j.emj.2016.02.001

[81] F. Brulhart, S. Gherra, and B. V. Quelin, "Do stakeholder orientation and environmental proactivity impact firm profitability?" *J. Bus. Ethics*, vol. 158, pp. 25–46, 2019.

[82] T. Donaldson and L. E. Preston, "The stakeholder theory of the corporation: Concepts, evidence and implications," *Corp. Its Stakeholders*, pp. 173–204, 1998. https://doi.org/10.3138/9781442673496-011