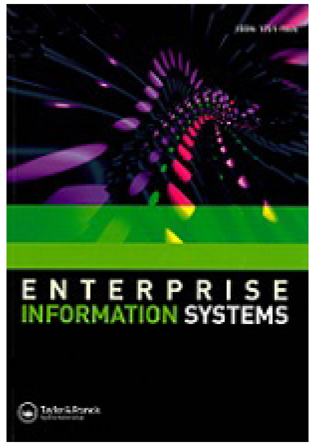

Customer Opinions Mining through Social Media: Insights from Sustainability Fraud Crisis - Volkswagen Emissions Scandal

Journal:	*Enterprise Information Systems*
Manuscript ID	TEIS-2021-0216.R2
Manuscript Type:	Research Article
Keywords:	Data Mining, Business Analytics, Customer Relationship Management (CRM), Enterprise Management, Enterprise Information Systems (EIS)

Customer Opinions Mining through Social Media: Insights from Sustainability Fraud Crisis - Volkswagen Emissions Scandal

Social media has emerged as a vital tool to understand customers and advance two-way communication between companies and customers. This paper uses social media data to investigate a well-known sustainability fraud, the Volkswagen emissions scandal (also called "emissionsgate" or "dieselgate") and focuses on the public opinions when the VW emissions scandal unfolded. The paper provides a comprehensive Tweet Analytic Framework for analysing tweets relating to this business ethics crisis, comprising three major analysis approaches: cluster analysis, sentiment analysis and time series analysis. The study involves a dataset of 29,764 collected tweets, which are separated into 2 stages - up to and after the peak point of public attention. They reveal the typical crisis development trend and peak point of the VW scandal, the strong condemnation and negative sentiment towards Volkswagen's fraud crisis, and the public concern regarding the topics of affected models and climate change. These findings can yield important insights for Volkswagen and practitioners to understand how customers' opinions changed, thereby managing sustainability fraud crises by improving the effectiveness of crisis management practices.

Keywords: emissions scandal; sustainability fraud; social media; cluster analysis; sentiment analysis; time series

1. Introduction

In an increasing number of companies, sustainability has become a mainstream issue that gives it a higher reputation and prominence (Steinmeier 2016). As companies further integrate sustainability into annual reports and have higher expectations for achieving sustainable performance, the pressure on them, as well as the financial interests, is rising. This can easily lead to increased opportunities for individuals and companies to participate in greenwashing and fraudulent behaviour in order to exploit sustainable efforts for their own advantage (Bartels, Moll, and Broekhof 2020). In other words, financial benefits not only bring positive motivation to improve sustainable performance but may drive towards the risk of fraud in related sustainable information.

In September 2015, the press reported that the US Environmental Protection Agency accused Volkswagen (hereafter referred to as "VW") of deploying and installing an illegal emission control device on 482,000 of its diesel-equipped vehicles sold in the United States. The defeat device resulted in intentional under-reporting of nitrogen oxide (NOx) emissions, enabling vehicles to comply with US federal emission standards and be certified when undergoing emissions tests (Topham 2015). VW subsequently admitted the allegation and revealed that the scandal's scope was 11 million vehicles sold worldwide rather than only the 482,000 vehicles sold in the United States. Based on Hartman (2015), the VW "sets the bar at a whole new level". The VW emissions scandal is not related to automobile quality or safety issues (as was the Aston Martin recall in 2014). Conversely, this VW emission scandal is a typical example of sustainability fraud, which is also seen as a new type of greenwashing behaviour (Siano et al. 2017). VW was cheating to meet the standard applied by the US emissions test. This involves sustainability deception to the public. Therefore, crisis management to rebuild public trust is key to recovery from the emissions scandal and for the sake of long-term survival. This can affect the development of automobile industry stakeholders, including companies of the whole industry and the government.

With the growing use of social media, many business crises and scandals have been impacted by consumer opinions expressed online. When the VW scandal occurred, both the media and individuals posted their viewpoints on the Volkswagen crisis on different platforms, especially online social platforms. In addition to gathering consumer opinions, social media platform also allows these to spread far and wide. Social media is an emerging tool for organizations to share information and update their interests, products, and services, which advances two-way communication between firms and customers (Poulis et al. 2019; Chun, Leem, and Suh 2021; Chung, Mustaine and Zeng 2021). During the crisis management stage, consumers use social media to share information regarding food safety, social crisis, environmental issues and even sudden emergencies (Gaspar et al. 2016; Ki and Nekmat 2014; Panagiotopoulos et al. 2016). Social media also helps companies discern what consumers care about most, how the involved company should respond, what management measures could be carried out, and how. Consequently, using data analysis methods to collect consumer responses towards business ethics crises as expressed on the Twitter platform, including opinions and sentiment, offers an efficient approach to crisis management. Indeed, social media has been a useful communication tool for crisis management, especially in cases of organizational crises involving food issues (Zhu, Anagondahalli, and Zhang 2017). Previous research has explored the expression of public opinions using social media platforms, while there has been little research on the use of social media data to analyze business ethics crises, especially with regard to sustainability fraud. To fill the gaps in social media research and provide useful insights into the VW sustainability fraud, this study focuses on the public opinions concerning this sustainability fraud during the months when the VW emissions scandal unfolded. Based on the research purpose, three research questions are proposed:

RQ1: What are customer sentiments regarding related tweets towards the Volkswagen emissions scandal?

RQ2: What are the most concerns and interests of customers regarding the Volkswagen emissions scandal?

RQ3: What are the important insights that companies and practitioners can know from the tweets analytical framework when a sustainability fraud crisis happens?

In order to address these research questions, we employed consumer opinion mining to understand the VW emission crisis on the Twitter platform. We captured and collected the tweets from 18/09/2015 to 1/12/2015. A total of 180,506 tweets were collected using the hashtags #volkswagenscandal, #vwscandal, #emissionsgate. After performing data pre-processing to a structured format, we gained a clean dataset that was ready for analysis. Compared with other studies that have discussed greenwashing and deception behaviour from the VW emission scandal (e.g. Lane 2016; Siano et al. 2017), this study proposes a comprehensive Twitter analytic framework for analyzing a large social media dataset. This study examines the collected tweets and reveals the strength and changes of scandal perceptions in a sample of tweets by applying data mining techniques combined with timer series analysis.

The remainder of the paper is organized as follows: section 2 comprises prior literature on business ethics and sustainability fraud, and a review of relevant recent social media research on crisis management. Section 3 and section 4 outlines the research methodology including research framework and data collection. Section 5 presents data analysis and results. In section 6 we discuss the implications and limitations of this study.

2. Literature Review

2.1 Business Ethics and Sustainability Fraud

With increased attention given to business ethics, scholars have increasingly shown interest in examining ethical issues at the organizational level (McLeod, Payne, and Evert 2016). Business ethics can be understood as a study of applied ethics or professional practices, which examines the ethical issues of productive and commercial activities that arise in the business environment (Moriarty 2016). Ethics applies to all aspects of business conduct and is related to the behaviours of individuals and entire organizations. In the contemporary business market, ethical issues involve diverse dimensions, such as honesty, professional conduct, environmental issues, integrity, and fraud. Brunk (2012) considered ethics as being socially responsible, which is often interrelated with CSR (Ferrell et al. 2019). Business ethics attract the scholars' attention as they affect decision-making, behaviours and performance (e.g. Payne et al. 2013; Shao et al. 2013). Corporate misconduct can have severe repercussions in the form of long-term harm to the company's reputation and image.

Business ethics is a huge field and vary significantly in the business field. There is also extensive literature focusing on sustainability topics (Quarshie, Salmi, and Leuschner 2016; Schuler et al. 2017). Nonetheless, limited research has focused on the unethical issue of the sustainable fraud crisis. Environmental protection is core to ethical business practices (Martinez and Bowen 2013). Brunk (2012) argued that "perceptions of being ethical is almost synonymous to abiding by the law." However, sustainability fraud is defined as fraud and misconduct committed either by sustainability or other professionals within an organization with respect to their sustainability-related work (Steinmeier 2016; Bartels, Moll, and Broekhof 2020). Incomplete or incorrect performance data may lead to more favourable economic benefits and is therefore fraudulent. Sustainability fraud can also take many forms, ranging from inappropriate variable compensation to false reporting. Sustainability fraud can be further understood by applying the Fraud Triangle, which generally entails: (1) the incentives and pressures to act fraudulently, (2) the perceived opportunity to commit fraud, and (3) the rationalization of fraudulent acts (Soltani 2014; Trompeter et al. 2013). As with many fraud risks, sustainability fraud can not only harm the company financially, but also threaten the public trust and confidence in companies and damage quality such as reputation, compliance with business ethics. They are seen as important drivers for promoting enterprises or their sustainable efforts (Bartels, Moll, and Broekhof 2020).

2.2 Social Media

For years, big data has emerged as one of the popular topics in business and other applicable fields (Sheng, Amankwah-Amoah, and Wang 2017; Salama, Kader, and Abdelwahab 2021). As a component of big data, social media data has also attracted much attention in their application in marketing (Ashley and Tuten 2015), education (Tess 2013), political events (Shirky 2011), stock market predictions and financial analysis (Bollen, Mao, and Zeng 2011; Eickhoff and Muntermann 2016; Liu et al. 2021).

Hansen, Shneiderman, and Smith (2010) defined social media as websites and online tools that strengthen users' interactions by exchanging information and sharing perceptions. Some social media applications, such as LinkedIn, Twitter, and Facebook, are the most frequently used social media by businesses, individuals and the government. However, social media tools have a diverse ecology, and they vary in terms of their nature, scope, and characteristics (Ngai, Tao, and Moon 2015). Social media has been an integral part of people's daily lives, including sentiment expression, personal activities posting and product review (Bian et al. 2016). As the public increasingly uses social media, social media has also become a powerful communication channel between enterprises and customers (Chan et al. 2016; Tajudeen, Jaafar, and Ainin 2018). The information conveyed by customers on their social networks plays an important role in influencing public opinion and behaviours. Many studies also demonstrated that accessing social media data can help businesses explore customer attitudes and satisfaction levels (Hajli 2013; Jin and Phua 2014; Ramanathan, Subramanian, and Parrott 2017) and build a brand reputation (Habibi, Laroche, and Richard 2014).

In this study, we focus on only one social media platform: Twitter. The text portions of Twitter, also called tweets, are usually related to events that involve many people in different parts of the world (Gaglio, Lo Re, and Morana 2016). Twitter allows brief posts of up to 280 characters together with images and other content (Cooper et al. 2022). This platform can provide real-time data for the public to express and share opinions and ideas. Meanwhile, Twitter hosts a large volume of opinions, which reflect online users' reactions to events or even social crises. It is also an appropriate source for opinion mining and sentiment popularity detection. Daniel, Neves, and Horta (2017) provided evidence that company event popularity is detected through sentiment analysis of tweets published by stakeholders in the Twitter universe. In addition, by using the collected tweets posted on Twitter, Ibrahim, Wang, and Bourne (2017) investigated the impact of online retailers' engagement with the online brand communities on users' perception of brand image and service.

Social media offers the opportunity to improve short-term and long-term understanding of consumers. Content mining from social media datasets is used to collect relevant content and identify real-time reflection (Thomaz et al. 2017). In the case of the VW emissions scandal, interested parties included not only the consumers directly affected, but also other car makers, potential consumers and the general public. All of these groups posted their opinions on Twitter or other social media platforms once the scandal information was exposed. The polarities and relative strengths of opinion, revealed with the help of clustering, sentiment and time series analysis, can help VW respond appropriately.

2.3 Social Media and Crisis Management

2.3.1 Crisis Management

Crises have become more frequent and are an increasingly familiar feature of the organizational landscape (James, Wooten, and Dushek 2011). Crisis management is defined as "a set of factors designed to combat crises and to lessen the actual damage inflicted by a crisis" (Coombs 2015). Thus, crisis management is the process by which organizational members avert or deal with emergencies to reduce the losses caused by the crisis. In the contemporary business environment, there is a commonly used three-stage approach describing crisis management, namely pre-crisis (precaution and crisis preparation), crisis (crisis recognition) and post-crisis (crisis recovery) (Coombs and Laufer 2018).

Crisis management is vital to manage, recover and even survive a crisis, not only for organizations but also for individuals and government, all of whom are stakeholders. Ki and Nekmat (2014) applied Situational Crisis Communication Theory (SCCT) and a sample comprising Fortune 500 companies to examine the effectiveness of crisis management using Facebook. Omilion-Hodges and McClain (2016) investigated the crisis lifecycle holistically within the university setting, from the effects of organizational channel selection to the way the organization's main stakeholders understand crisis information.

2.3.2 Crisis Management in Social Media Research

Social media platforms are taking the place of traditional media resources and become a vital means to report an event (Ma et al. 2021). Hence, the public increasingly relies on mobile and social media technologies during crises and other unanticipated events (Lachlan et al. 2016). In other words, when crises occur, the public tends to use social media sites such as Twitter and Facebook to obtain real-time information and share opinions about events as they unfold (Pew Research Center 2019). In many cases, social media platforms can serve as first-hand messages source, enabling locals to report immediate information after the crisis broke out. The research of Spence, Lachlan, and Rainear (2016) proved that using social media to collect data on crises and disasters is great practice, with high utility. Nevertheless, the use of social media tools in crisis communication, and social media's role in managing crises are developing in a relatively unexplored context (Graham, Avery, and Park 2015).

Twitter is emerging as one of the dominant social reporting tools for disseminating social crisis information, promoting online public communities as first responders who can collectively cope with social crises (Oh, Agrawal, and Rao 2013). Consequently, the literature is paying increasing attention to crisis management using social media such as Twitter. A growing body of literature studies social media together with crisis management, such as Hurricane Sandy (Kogan, Palen, and Anderson 2015), the Ebola virus (Kim et al. 2016), Airlines' overbooking crisis (Ma et al. 2019), and the SARS-CoV-2 pandemic (Choi et al. 2020). Gao et al. (2020) evaluated the prevalence of mental health problems and examined their relationship with social media exposure during the COVID-19 outbreak. Besides, the sentiment analysis technique is also employed to analyze the response of social media users to unexpected and potentially stressful social events, including social crises (Gaspar et al. 2016). The polarity and strength of online users' opinions tend to reflect the crisis development trend. Ji et al. (2015) explored the sentiments of tweets related to public health concerns: clinical science, epidemiology and mental health. These have shown the possibilities for the public or organizations to come together in creating responses to events.

Although researchers have extensively analyzed social media data in different contexts, the areas of business crisis management are still under-researched. In this study, we focused on analyzing the VW emissions scandal crisis using Twitter-generated messages. Following Terpstra et al. (2012), who explored the possibilities of real-time and automated analysis of Twitter messages during the crisis, we collected tweets from the beginning of the emissions scandal outbreak and throughout the whole lifecycle of the crisis. Considering the demands and attitudes of consumers is a critical step in designing an effective communication strategy; we also considered the entire lifespan of the company's crisis.

3. Research Framework

Figure 1 illustrates a Tweets Analysis Framework for analyzing tweets related to the Volkswagen emission crisis. Firstly, we started with the data collection process. Similar to other social media research (Zhao, Zhan, Jie 2018; Ibrahim and Wang 2019; Abd-Alrazaq et al. 2020), the following analysis is implemented after pre-processing the tweets, aiming to filter out and remove the noise of social media data, namely converting unstructured data to a structured format. Next, using the peak point model, the dataset was separated into Stage 1 (18/09/2015 to 23/09/2015) and Stage 2 (24/09/2015 to 1/12/2015).

Following the preparation of the dataset, we employed some text-mining techniques to analyse the social media data for each stage, such as cluster analysis and sentiment analysis. Specifically, we utilised SentiStrength, a lexicon-based sentiment detection program, to analyse tweet sentiment. This tool makes use of a dictionary of positive and negative words (Daniel, Neves, and Horta 2017), and it estimates two sentiment strengths: -1 (not negative) to -5 (extremely negative) and 1 (not positive) to 5 (extremely positive). Its algorithm adopts linguistic rules (e.g. emoticon and negations) when computing positive and negative outputs (Thelwall, Buckley, Paltoglou 2012).

We also applied QDAminer software to carry out the cluster analysis and multidimensional scaling (MDS) analysis, which is a vital tool for identifying potential topics in a discussion group. Based on cluster analysis results, the co-occurrences of trending keywords are evaluated and grouped. According to Tse et al. (2016), we used Jaccard's coefficient as the index of similarity of co-occurrence to form impression groups. The Jaccard's coefficient was given by:

$$\text{Jaccard's coefficient} = [a / (a + b + c)]$$

Where "a" represents the number of occurrences in both items (i.e. the number of matching ones); "b" and "c" represent the case where one item is found but not the other.

Next, by using MDS, we created a positioning map of the different clustered groups and displayed their strength of the relationship in a two- or three-dimensional space (Tse et al. 2018; Ma et al. 2021). During this process, we also combined the time series analysis with text-mining results. Ibrahim and Wang (2019) considered that understanding time series is essential to uncover the hidden trends and insights in exploring a set of longitudinal data. Time series analysis was conducted to divide the shorter time ranges using the captured daily tweets, which can help explore sentiment changes over time and identify the peaks of Twitter activities to understand underlying patterns at a particular time (Ranco et al. 2015). The results are reported in Section 5.

[Insert Figure 1 here]

4. Data Collection and Pre-processing

The dataset used in this study comprises the Twitter posts directly associated with the Volkswagen emission scandal crisis (i.e. "UKVolkswagen", "Volkswagenscandal", "VWscandal", and "VWgate") during the period from 18/09/2015 to 1/12/2015. A total of 180,506 tweets were collected as shown in Figure 1. These include tweets and retweets generated both by individual online users and by Volkswagen group official accounts such as @Volkswagen_UK, @vwgroup_en, and @Volkswagen_USA. The tweets were purchased from GNIP, the Twitter data official provider. Due to the nature of the social media data that is unstructured, noisy, informal and large in volume, data pre-processing is a crucial step before starting the social media data analysis (Zhao, Zhan, and Jie 2018; Stieglitz et al. 2018; Abd-Alrazaq et al. 2020). In the cleaning process, we conducted the following procedures. We used keyword searches to filter the tweets with noise and redundancy, which included the hashtag (#) and mention (@). Only English language tweets were selected, in order to remove complications that might arise from analyzing multilingual tweets (Thelwall, Buckley, and Paltoglou 2011). As our data was crawled according to certain search criteria, we cut out the associated high-frequency words in the body of comments; for example, "UKVolkswagen", "Volkswagenscandal", "VWscandal", "VWgate", "Volkswagen", "scandal", "VW", and "RT" were substituted by blanks. To ensure the collection of valid customer opinions, we also filtered out 148,964 tweets containing a URL. We excluded these messages since they only involved the content from the news or the retweet of a news article, however, they did not include any comments from the customers. This left 31,542 tweets without URL, of which 1,778 were invalid tweets. Finally, as Fig. 1 shows, we removed noise and data redundancy to obtain 29,764 tweets of pure comments in the final dataset.

With this dataset, we began categorizing the types of tweets and hashtags by applying basic statistical analysis. Among the 180,506 original tweets, about 30% were generated by a web client, and nearly 70% were generated by mobile terminal clients.

Besides, we applied the QDA miner software to analyze the hashtags statistics. This paper identifies 29,764 unique hashtags in the tweets dataset. 29,764 tweets include two or more hashtags. For example, a single tweet may contain the hashtags #Volkswagenscandal, #reputational risk and #emissions. This indicates that a large number of tweets intersect with multiple areas of interest (Chae 2015).

5. Data Analysis

5.1 Data Separation

The pivotal event was the resignation of VW CEO, Martin Winterkorn on September 23, 2015. Figure 2 shows the time trend of the volume of tweets. The number of customer comments reached its first and highest peak on that date. This is the climactic point of public attention, revealing 23 September was a key turning point of this corporate crisis, with the resignation of the CEO as the critical event. It is for this reason that we split the data into two stages: up to, and after 23 September.

[Insert Figure 2 here]

To get more accurate knowledge of what the customers were concentrating on, the dataset was separated into 2 stages, namely Stage 1 (18/09/2015 – 23/09/2015) and Stage 2 (24/09/2015 – 1/12/2015) according to the peak point by panels (also called peak point of public attention). The clustering analysis for these two stages is discussed in sections 5.2 and 5.3 respectively. Clustering analysis can help automatically organise and explore information from unstructured text data (Liau and Tan 2014).

Segment data analysis allows more attention to detail, and thus yields more accurate analysis. If the analysis is applied only to the data as a whole, some key points could be missed or misunderstood, thereby hindering managers from making correct and timely decisions on how to respond to a crisis, and therefore leading to further deterioration of the situation. For instance, looking only at the entire data, managers may find that one tweet from the Wall Street Journal was retweeted many times, and then pay more attention to it. In fact, we put considerable effort towards processing the tweets from the Wall Street Journal. However, the results of the segment data analysis revealed that the main issues for online users were climate change, air pollution and VW's use of fake data. The two stages of data analysis are presented as follows.

5.2 Clustering Analysis for Dataset of Stage 1

5.2.1 Word frequency and distribution discovery for dataset of stage 1

Table 1 reports the word frequency findings for stage 1. As can be seen from the table, "car", "emission" and "company" are the top 3 most frequently occurring words. These are all directly related to the "emission state", indicating that it was the involved cars and their emission problems that were discussed most. Unexpectedly, the words "Hitler" and "Adolf" also occurred with high frequency. Detailed analysis of dendrograms revealed a dendrogram of a Hitler Jokes group. This can be traced back to the history of the VW group, which was founded on the idea of "the people's car" proposed by Adolf Hitler, an idea that cars should be of the people and for the people. This scandal certainly runs contrary to that idea.

[Insert Table 1 here]

5.2.2 Mapping for the dataset of stage 1

As can be seen from Figure 3, the 2D MDS map reveals the mapping of the stage 1 dataset. Five topic groups related to the Volkswagen emissions scandal are illustrated by the clustered keywords. The circles represent the clustered main keywords of the dataset. The closer the circles are, the higher the tendency for co-occurrence, and vice versa.

[Insert Figure 3 here]

5.2.3 Dendrogram mining for dataset of stage 1

Figure 4 to Figure 8 demonstrate the dendrogram mining results for the stage 1 dataset.

[Insert Figure 4 here]

[Insert Figure 5 here]

[Insert Figure 6 here]

[Insert Figure 7 here]

[Insert Figure 8 here]

• Climate Change

Figure 4 shows the dataset dendrogram of the Climate Change group, mined from tweets expressing customer opinions. Customers were most concerned about the climate change implications of the on-road NOx emissions being 40 times higher than those found in the test lab. Representative comments include:

"VW scandal neatly shows how climate change is a product of unrestrained capitalism and wealth accumulation. Climate change is a class issue."

"@HillaryClinton VW scandal harmful for EU credibility on climate change talks. Big gap between the high level claims and the actual evidence."

• Affected Models

In Figure 5, the dendrogram of the Affected Models group, relates to the models involved in this emissions scandal, such as Jetta, Golf, and Passat. Owners of these cars felt disappointed and cheated. The dataset includes comments such as:

"You've disappointed n cheated a lifelong VW owner. My 2010 Jetta SportWagen TDI was my pride n joy. Why?"

"@VW I'm so disappointed in Volkswagon. I bought my Passat TDI because I thought this company had integrity. Never again. #DumpUrVeeDub."

• Fake Data

As revealed in Figure 6, the dendrogram of the Fake Data group, the public expressed anger and discontent about the falsification of data. The dataset includes tweets such as:

"RT @Robert___Harris: After Volkswagen scandal, Germans may have to give the Greeks rather fewer lectures on falsifying data."

"The Volkswagen Scandal Will Hurt the Auto Industry More Than You Think: The whole issue of fudging data to meet emission norms is not…"

• Hitler Jokes

As shown in Figure 7, the dendrogram of the Hitler Jokes group, the dataset contained many references to "Adolf Hitler". This was in the context of jokes reflecting

VW's use of software to cheat in the EPA test. The dataset reveals that one Twitter user, named DaveSFoley, was an opinion leader with regard to this kind of joke.

"RT @DaveSFoley: Volkswagen scandal is a dark stain on the legacy of the company's founder, Adolf Hitler."

"@Volkswagen Congratulations on your attempt to break American law. Your founder, Herr Hitler, would approve. #despicable."

The above comments from the dataset show that customers were using jokes to criticize VW for its cheating to pass the EPA test. They also revealed that online users prefer to express their blame in a joking manner (Tse et al. 2016).

- Blaming Diesel

For years, Europe has been the main market for diesel vehicles and diesel has been considered the cleanest fuel. However, as shown in Figure 8, the VW sustainability fraud incident exposes the public that Diesel engine emissions pollute the environment. Many customers blamed diesel as a dirty fuel and accused VW of dishonesty via Twitter. Representative comments include:

"#Volkswagen scandal just more proof that diesel is a horrible dirty fuel. ""Clean Diesel" an oxymoron."

"RT @breakingauto: BREAKING: Diesel @VW scandal gives many auto writers their first-ever chance to write something negative about a car company."

5.3 Clustering Analysis for Dataset of Stage 2

5.3.1 Word frequency and distribution discovery for dataset of stage 2

Table 2 presents the word frequency for the second stage of the dataset, which covers the period 24/09/2015 to 1/12/2015. The words "car", "emission", and "software" are highly related to the search criteria terms. The words "auto", "German", "diesel", "TDI" and "Audi" are also the focus in this stage. This can be explained by the fact that it was during this period that the public learned which models were more affected. The development of the crisis can be traced through the frequency with which words occurred in the tweets, and different topics emerged.

[Insert Table 2 here]

5.3.2 Mapping for the dataset of stage 2

Figure 9, below, shows the mapping of the dataset in stage 2. It reveals four topic groups related to the Volkswagen emissions scandal, namely "Risk Spreading", "Affected Models", "Golf Involved" and "Air Pollution".

[Insert Figure 9 here]

5.3.3 Dendrogram mining for dataset of stage 2

Dendrogram mining for the dataset of stage 2 revealed four groups, as illustrated from Figure 10 to Figure 13. At this stage consumer tweets are more related to the details of the emissions scandal, including the risk spreading, affected models, the Golf model, and worries about air pollution.

[Insert Figure 10 here]

[Insert Figure 11 here]

[Insert Figure 12 here]

[Insert Figure 13 here]

• Risk Spreading

Figure 10, the dendrogram of the Risk Spreading group, includes the words "worldwide" and "million", reflecting the fact that millions of cars worldwide were affected. It is an unbelievable number. From the initial unwillingness to admit to the scandal, to the exposure of just how many cars were equipped with cheating software, the crisis had spread rapidly, and this can also be seen from original tweets such as:

"RT @AFP: #BREAKING VW Scandal: Audi says 2.1 million cars worldwide fitted with emission-cheating software."

"#VolkswagenScandal: 21.4 lk @Audi cars globally affected by emission scandal."

• Affected Models

As in the case of stage 1, the dendrogram of the Affected Models group is also shown in Figure 11. This is one of the representative responses from customers towards the VW sustainability fraud incident. Many VW TDI owners shared their stories and expressed their emotions on Twitter. Some of them have a strong attitude to change car brands after the VW emission scandal. In other words, many previous customers' purchase intentions are changing, which is a big blow and punishment for Volkswagen. The dataset includes tweets such as:

"Happy to hear @LeoDiCaprio is doing a movie on #VolkswagenScandal. On behalf of the VW TDI owners, I hope u capture some of our stories."

“RT @jen2seely: @VW I loved my Passat TDI so much I said if I won the lotto I'd still drive it. It's my 3rd VW. #heartbrokenfan #nowwhat”

• Golf Involved

VW admitted in official tweets that the Golf, Jetta, and A3 were among the models involved in the emissions scandal. As shown in Figure 12, the dataset results reflect the previous popularity of the Golf model, and customer reaction at its involvement in “emissionsgate”. The public expressed their opinions in tweets such as:

“@vwgroup_en @UKVolkswagen @VWUKHelp @IanJPlummer VIN #WVWZZZ1KZCM639035 compensate & replace with white Golf 1.0 TSI DSG SE Bluemotion Estat5.”

“Very disappointed to find out that my #VWGolf is one of those affected by #vwscandal and that I've had to find out myself. #NotGoodEnough.”

• Air Pollution

Figure 13 reflects the public concern over air pollution caused by vehicle emissions. Tweets included:

“RT @ConversationUK: If some good can come out of #VolkswagenScandal it's that public will be more clued up about air pollution from cars CleanAirLondon.”

“WHAT ABOUT LIVES? 29,000 deaths from vehicle air pollution last year!”

5.4 Time Series Analysis

In Figure 14, a time-series diagram includes tweet frequency, word clouds and sentimental analysis. According to Vergeer and Franses (2016), continuous data collection across time allows dynamic analysis, such as time series analysis. In this study, the time series concept is applied to divide the observation period into time intervals using the captured daily tweets. This section separates the data into two subsets of data during the research period in order to examine the changes in customer opinions as the emission scandal developed.

This section is to further investigate the tweet distribution of the recall scandal over time, a 10-day timeline is added to the above sentiment classifier, and a time series analysis is employed to compare the numbers of the tweets and their sentiment scores captured in different time. The 10-day period (23 February to 3 March 2016) is broken down into a half-day manner (am/pm) to study the variations in the popular topics and sentiment. Hence, the original dataset is separated into 20 sub-datasets for the time series analysis in Figure 6.

[Insert Figure 14 here]

Based on this time series analysis, we can conclude the following five points:

(1) Typical crisis development trend

As shown in Figure 15, González-Herrero and Pratt (1996) illustrated how crises follow a sequential path along with four phases: birth, growth, maturity and decline. They divided the crisis into identifiable stages and explained how the crisis has changed over time. This basic model provides a simple and effective description of the crisis life cycle. This study refers to Renn (1991) and Panagiotopoulos et al. (2016)'s categories of message components to classify tweets when assigning captured tweets. We provided more complete stories unfolding from the VW emissions scandal on a time series to examine the frequency of tweets and how the volume of information corresponds to the nature of the actual risk events. The number of tweets is expected to fluctuate depending on the intensity of the crisis. In 2014, the California Air Resources Board (CARB) received a study published by the International Council on Clean Transportation (ICCT) that showed significant differences in nitrogen levels for VW diesel vehicles between bench testing and road operation (Reuters 2016). VW's initial reaction was just to give a technical response. This remained the company's position until 18 Sept 2015. From this point, the emissions scandal followed a typical crisis lifecycle.

[Insert Figure 15 here]

Birth stage: On 18/09/2015, the EPA issued a public notice of violation of the Clean Air Act to VW, alleging that the model year 2009-2015 VW and Audi diesel cars with 2.0-litre engines included defeat devices. The "emissionsgate" scandal broke out.

Growth stage: From 18/09/2015 to 23/09/2015, the crisis spread, and an increasing number of customers expressed their opinions about the scandal through social media.

Maturity stage: On 23/09/2015, the CEO left the company. The crisis widened, as it was admitted that the cheating software had been installed in diesel vehicles sold across Europe.

Decline stage: From 24/09/2015 to 1/12/2015 (the dataset is from 18/09 to 1/12/2015), with the CEO's resignation and other responses to this scandal, "emissionsgate" declined.

(2) Resignation of the former CEO as the peak point of the scandal

"Emissionsgate" triggered a butterfly effect; from the breakout of the crisis on 18 September 2015, events moved quickly. On the evening of September 23, the German Volkswagen CEO Martin Winterkorn announced his resignation.

In fact, results of the time series analysis show that the resignation of the CEO was the turning point of the whole scandal. Before this point, the increase in the frequency shows that the crisis was growing rapidly as a shocked public responded to VW's admission that 11 million cars worldwide had been fitted with the defeat devices. Following Martin Winterkorn's decision to quit, the frequency trend began to decline.

Results for the second stage section of the dataset reveal several peaks coinciding with particular events. On 28 September 2015, German prosecutors launched an investigation of Winterkorn. On 30 September, it became known that almost 1.2 million VW diesel vehicles in the UK were affected. On 8 October, police raided VW's headquarters and on the same day VW's US boss, Michael Horn, said that he felt personally deceived. On 13 October, it was announced that Leonardo DiCaprio's production company wanted to make a film about the scandal.

When talking about business ethics, Francis and Armstrong (2003) argued that ethics are not easily achieved in the turbulent and competitive environment of the business world. Only genuine commitment can make a difference. The resignation of the former CEO indicated the company's commitment to addressing this scandal.

(3) Sentiment trend matches the frequency trend

The upper part of Figure 14 shows the frequency trend over time, while the lower part shows how the sentiment strength varies over the same period. The figure indicates that the frequency of tweets is related to the strength of sentiment. Higher frequency is associated with stronger sentiment.

Sentiment analysis can help the company understand emotional polarity by detecting the sentiment of tweets. It is worth noting that the two curves in Figure 14 are largely symmetrical. They reflect peaks in the level of negative emotion. Indeed, during this crisis, all the comments from consumers can be classified as highly negative, often using more than one negative word to express emotion. Especially from the perspective of air pollution and trust, many tweets were found to have an extremely negative sentiment towards the VW scandal. With the first three peaks, emotions are deepening, while from the fourth peak sentiment begins to weaken. From this point, according to sentiment trend analysis of the entire dataset, the negative emotion subsides, and the whole crisis is now under control.

(4) Strong condemnation of VW's cheating

The word clouds of the tipping points show that at the beginning of the crisis, especially before 30 September 2015, the word "cheat" occurred with high frequency. In the entire tweets dataset, 1,602 messages contain the words "cheat", "cheating" or "cheated". The element of cheating has led to a loss of trust in the VW group. In the case of a simple problem, for example with auto parts, scandals can be overcome within a few days, with a headline apology and recall, but VW's behaviour has linked the company with the issues of environmental air pollution, cheating and other business issues. This is much more serious, and it will be difficult to regain consumer trust. Indeed, public shock at the behaviour of a previously trusted German enterprise may mean that the national image of Germany, and positive associations of "Made in Germany" have been damaged by this scandal.

It is clear from the discussion around "cheating" that the public attaches great importance to the performance of the ethical level. They also output more negative comments on corporate behaviours such as cheating. By using cheating software to pass an environmental test, Volkswagen violated normal standards of business ethics in two ways: in terms of company honesty and with regard to the environment.

(5) High concentration on air pollution and climate change

The word clouds also reveal that, during the sustainability fraud scandal, customers were particularly concerned about air pollution and climate change. A study published in *Nature* shows that diesel vehicles emit far more NOx under real-world operating conditions than during laboratory certification tests (Anenberg et al. 2017). Nitrogen oxide is a key contributor to outdoor air pollution. In this VW emission scandal, NOx emissions are a threat to public health as they are pollutants that can lead to more serious respiratory diseases and aggravating heart and lung disease (Mathiesen and Neslen 2015). Long-term exposure to pollution would exacerbate mortality.

However, despite the fact that the emission of more NOx is directly damaging to health, consumers' concerns were focused more on the environment and climate change. This can be explained by the fact that, while the public considers that higher on-road emissions lead directly to pollution and climate change, they do not have much knowledge about the real environmental effects of NOx emitted by diesel engines. From the data results analysis, it seems that the public was not aware of the personal effects of NOx. They considered themselves to be bystanders, not victims.

6. Discussions and Conclusions

The emergence of the VW emission scandal aroused widespread discussion among the public on social platforms. It is a new form of automobile reputational crisis, which is also characterized instead as a form of business ethics crisis. This study focuses on customers' perceptions of and responses to the sustainability fraud crisis through their opinions expressed on the social media microblogging platform, Twitter. To achieve this, we investigated the Volkswagen emission scandal using a comprehensive framework for analyzing tweets. Within this framework, a combination of data mining techniques, including cluster analysis, sentiment analysis and time series analysis, was used to determine the topics of customers' concern and the trend of sentiment.

Several insights emerged from this study as an attempt to explore the users' engagement on the Twitter platform in response to the VW emission scandal. Specifically, we began with a data separation based on the volume of tweets. This was followed by the MDS analysis with the integration of cluster analysis, aiming to identify what the users are concerned about in two stages. Based on the MDS clustering approach, nine clustered groups were identified. They reflect the timely responses and thoughts of customers on the VW emission scandal during the crisis. VW and its practitioners can identify and explore emergent segments of customers' concerns. For instance, the clustered groups of "affected models" appeared at two stages in the time series. This represents the high frequency and importance of this topic during the crisis, with many car models and their owners affected by the Volkswagen emissions crisis. Next, we performed a time series analysis of the trends exhibited by the volume and sentiment of tweets on the time series. According to Ibrahim and Wang (2019), sentiment analysis was designed to provide insights to understand the opinions and trends of engagement activities on the social platform. The time-series diagram also includes word clouds and clarifies that the VW emissions scandal followed a typical crisis lifecycle. The result indicates that the sentiment of users towards the VW emission scandal was negative. During the crisis period, the higher the frequency of tweets, the stronger the negative emotions of users. The public had strong condemnation of VW's fraudulent behaviour and their concerns about air quality and the environment via Twitter.

We emphasize the importance of users' opinions on social media, especially during emergencies or crises. This approach can use collective intelligence and encourage public participation in making better and timely decisions in response to the development trend of the company's crisis. The proposed framework also helps measure the effectiveness of crisis management in the post-crisis period. The VW sustainability fraud crisis involves the interests of the public and different organizations, and its potential consequences need to be emphasized. As mentioned in the result analysis, a number of diesel vehicles sold by VW were affected. Volkswagen plays a crucial role in Germany's automobile industry (Bach 2015). The VW emission scandal has not only aroused public concern about Volkswagen vehicles but may also raise questions about the supply chain for related diesel vehicles. In fact, there has been an increasing importance of sustainability in business strategy and corporate reputation. This scandal has seriously threatened the reputation of the entire automotive industry and the efforts to achieve environmental sustainability using technologies.

To summarize, the ethics of business conduct is increasingly under public scrutiny. We highlighted a case of fraud committed with sustainability information. It is unethical and illegal for Volkswagen to falsify green performance data to enhance its compliant and sustainable positioning. In addition to damaging the brand reputation of Volkswagen and causing financial losses to the company (e.g. government fines and loss of sales), this may also harm public health and have a knock-on effect throughout the automotive industry. All these, therefore, come not only with a growing concern for the future of the planet, but also with financial interests, and an understanding of the intertwined between the two.

6.1 Research Contributions and Implications

This paper contributes to the literature in several ways. Firstly, it adds to the stream of literature examining sustainability fraud behaviour in the business ethics context. Secondly, previous studies strongly advocate the use of social media as a communication tool. They focus on the benefit of social media usage to communicate instantly when crises break out (Panagiotopoulos et al. 2016; Zhu, Anagondahalli, and Zhang 2017). Some researchers focus on the customer opinion and sentiment around brands (Jin and Phua 2014), marketing (Daniel, Neves, and Horta 2017), service behaviour and value creation (Agnihotri et al. 2012), while others analyze data from social media in different contexts, including supply chain practices (Chae 2015) and food fraud crises (Tse et al. 2016) in order to yield findings that can provide the industry with effective strategies in marketing and crisis communication. This paper also proves that social media data can be a valuable source for understanding the opinions of online users. It shows how companies use social media data to better understand the public's perceptions of business ethics and to improve companies' understanding of customers, especially during times of sustainability fraud crisis. This can further enhance our knowledge of the connection between business ethics and social media and contribute to social media research. Thirdly, this study proposes a comprehensive Twitter analytic framework to analyze the customers' opinions by collecting massive social media data. Meanwhile, in contrast with previous studies that only focused on one or two techniques (e.g. Kim et al. 2016), we extend the existing literature by presenting multiple data mining techniques in a social media context, which can help improve the quality of analyzing social media data. Cluster analysis and MDS analysis are performed to explore the core topics and knowledge of social networking (Shiau et al. 2017). We also investigate users' responses towards VW sustainability fraud, including the changes in users' opinions and sentiment during the time series.

As McLeod, Payne, and Evert (2016) proposed, it is difficult to ascertain and measure ethics as they are abstract, evolving, and inherently nested in all levels of analysis. In terms of business ethics, the perceptions of ethics vary greatly in different cultures and business disciplines. Hence, we need to understand ethics and its diverse forms according to the characteristics of different companies and industries and make appropriate decisions to achieve positive development on the ethical level. Other researchers and scholars can use the design and findings of this study as a case to continue the theoretical contributions of the field for many years to come.

This paper also yields important practical implications. The VW emissions scandal has revealed a significant challenge in managing responses to business ethics on a social media platform. Corporate fraud scandals have a significant impact on firms and customers. The study of the UK horsemeat scandal by Tse et al. (2016) also provided an analysis case. In addition, environmental vulnerability and topics related to environmental sustainability have always been topics of global concern. Hence, managerial implications can be derived from that study as our results provide companies and managers with deeper insights into sustainability fraud crises through scrutinizing public attitudes and sentiments about the VW emissions scandal. It brings to light the importance of a clear commitment from executives and managers to sustainability to prevent sustainability fraud. It is also vital for practitioners to understand potential sustainability fraud risks they should be aware of and the actions they can take to prevent and act on it, particularly in companies with global brands, where the risk to the business can be perceived to be greater.

Furthermore, our research is of practical value to companies as we provided an updated social media analysis technique for crisis management. In a competitive market, social data analytics are increasingly important for modern enterprises. Firms need to pay more attention to social media data as user opinions expressed through social media can help companies understand public attitudes towards the crisis while retaining the company's brand loyalty and customer stickiness. A multiple text mining approach was conducted in this study to give companies a dynamic view of how customer opinions evolve over time. The frequency and weight of the topics in the time series can provide insight into the areas that should be prioritized. We also applied real-time Twitter data to detect the changes in users' sentiment in response to the VW emission scandal. Such information enables firms to observe optimal timing for the execution of specific strategies and make corresponding adjustments to their operations and strategies more promptly, which is in accordance with Poulis et al. (2019). When companies are able to grasp public sentiment, they can take targeted actions to weaken the strength of negative opinions, making it easier to control the development of the crisis. The recovery cycle of the corporate crisis will also be shortened, and the company will more easily recover to its pre-crisis state. Simultaneously, companies can monitor other services and design improvement strategies, thereby facilitating customer relationship management and market forecasting.

6.2 Limitations and Future Research

Firstly, there is still limited research on sustainability fraud, which calls for more research, such as assessing the potential determinants and preventive measures of sustainability fraud. Another limitation and associated research opportunity related to the sample size. For instance, this research only focuses on customers who are accustomed to sharing their opinions on social media. Future research may combine surveys or interviews methods to examine other groups of customers to achieve a more accurate representation of the population. This research collected the tweets using certain selected search terms (i.e. #volkswagenscandal, #vwscandal, #emissionsgate). It is suggested that future research can use keywords such as "Volkswagen" (or any of its variants) and more related search terms to gain a more comprehensive sample. Additionally, notwithstanding the text-mining analysis can effectively identify meaningful topics from textual data, this research applied an unsupervised approach, it is challenging that the obtained clustered topics might be distracted by the noises that exist in the social media data. For example, some extracted topics in this study are not closely related to the VW sustainability fraud (e.g. AMP). Hence, we suggest that future researchers consider engaging in more advanced methods to reduce noise when addressing social media data, such as machine learning assist text-mining or computer aid text-mining. Moreover, the framework proposed in this paper to discover potential information for crisis management was designed specifically for Twitter. It could be extended to different social media platforms and thus provide companies and practitioners with more perspectives and insights. Future researchers may consider using different social media platforms with longitudinal data to validate the proposed framework in this study.

Reference

Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M. and Shah, Z. 2020. "Top concerns of tweeters during the COVID-19 pandemic: infoveillance study." *Journal of Medical Internet Research* 22(4): e19016.

Agnihotri, R., P. Kothandaraman, R. Kashyap, and R. Singh. 2012. "Bringing "social" into sales: The impact of salespeople's social media use on service behaviors and value creation." *Journal of Personal Selling & Sales Management 32*(3): 333-348.

Anenberg, S. C., J. Miller, R. Minjares, L. Du, D. K. Henze, F. Lacey, C. S. Malley, L. Emberson, V. Franco, Z. Klimont, and C. Heyes. 2017. "Impacts and mitigation of excess diesel-related NOx emissions in 11 major vehicle markets." *Nature* 545(7655): 467-471.

Ashley, C., and T. Tuten. 2015. "Creative Strategies in Social Media Marketing: An Exploratory Study of Branded Social Content and Consumer Engagement." *Psychology & Marketing* 32(1): 15-27.

Bach, D. 2015. "Seven reasons Volkswagen is worse than Enron." *The Financial Times,* September 27. Accessed 28 November 2020. https://www.ft.com/content/cf9f73e8-62d6-11e5-9846-de406ccb37f2

Bartels, W., V. Moll, and M. Broekhof. 2020. "The rising challenge of sustainability fraud." *The KPMG*, 2020. Accessed 13 May 2021. https://assets.kpmg/content/dam/kpmg/nl/pdf/2020/services/the-rising-challenge-of-fraud.pdf

Bian, J., K. Yoshigoe, A. Hicks, J. Yuan, Z. He, M. Xie, Y. Guo, M. Prosperi, R. Salloum, and F. Modave. 2016. "Mining Twitter to assess the public perception of the "Internet of Things"." *PloS one* 11(7): e0158450.

Bollen, J., H. Mao, and X. Zeng. 2011. "Twitter mood predicts the stock market." *Journal of Computational Science* 2(1): 1-8.

Chae, B. 2015. "Insights from hashtag #supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research." *International Journal of Production Economics* 165: 247-259.

Chan, H. K., X. Wang, E. Lacka, and M. Zhang. 2016. "A mixed-method approach to extracting the value of social media data." *Production and Operations Management* 25(3): 568-583.

Choi, H., W. Cho, M. H. Kim, and J. Y. Hur. 2020. "Public health emergency and crisis management: case study of SARS-CoV-2 outbreak." *International Journal of Environmental Research and Public Health* 17(11): 3984.

Chun, H., B. H. Leem, and H. Suh. 2021. "Using text analytics to measure an effect of topics and sentiments on social-media engagement: focusing on Facebook fan page of Toyota." *International Journal of Engineering Business Management* 13: 1-11.

Chung, W., E. Mustaine, and D. Zeng. 2021. "A computational framework for social-media-based business analytics and knowledge creation: empirical studies of CyTraSS." *Enterprise Information Systems* 15(10): 1460-1482.

Coombs, W. T. 2015. *Ongoing Crisis Communication.* 4th ed. Thousand Oaks, CA: Sage.

Coombs, W. T., and D. Laufer. 2018. "Global crisis management-current research and future directions." *Journal of International Management* 24(3): 199-203.

Cooper, K., O. Dedehayir, C. Riverola, S Harrington. and E. Alpert. 2022. "Exploring consumer perceptions of the value proposition embedded in vegan food products using text analytics." *Sustainability* 14(4): 2075.

Daniel, M., R. F. Neves, and N. Horta. 2017. "Company event popularity for financial markets using Twitter and sentiment analysis." *Expert Systems with Applications* 71: 111-124.

Eickhoff, M., and J. Muntermann. 2016. "Stock analysts vs. the crowd: Mutual prediction and the drivers of crowd wisdom." *Information & Management* 53(7): 835-845.

Ferrell, O. C., D. E. Harrison, L. Ferrell, and J. F. Hair. 2019. "Business ethics, corporate social responsibility, and brand attitudes: An exploratory study." *Journal of Business Research* 95: 491-501.

Francis, R., and A. Armstrong. 2003. "Ethics as a Risk Management Strategy: The Australian Experience." *Journal of Business Ethics* 45(4): 375-385.

Gaglio, S., G. L. Re, and M. Morana. 2016. "A framework for real-time Twitter data analysis." *Computer Communications* 73: 236-242.

Gao, J., P. Zheng, Y. Jia, H. Chen, Y. Mao, S. Chen, ... and J. Dai. 2020. "Mental health problems and social media exposure during COVID-19 outbreak." *Plos one* 15(4): e0231924.

Gaspar, R., C. Pedro, P. Panagiotopoulos, and B. Seibt. 2016. "Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events." *Computers in Human Behavior* 56:179-191.

González-Herrero, A., and C. B. Pratt. 1996. "An integrated symmetrical model for crisis-communication management." *Journal of Public Relations Research* 8(2): 79-105.

Graham, M. W., E. J. Avery, and S. Park. 2015. "The role of social media in local government crisis communications." *Public Relations Review* 41(3): 386-394.

Habibi, M. R., M. Laroche, and M. O. Richard. 2014. "The roles of brand community and community engagement in building brand trust on social media." *Computers in Human Behavior* 37: 152-161.

Hajli, M. N. 2013. "A study of the impact of social media on consumers." *International Jouranal of Market Research* 56(3): 387-404.

Hansen, D., B. Shneiderman, and M. A. Smith. 2010. *Analyzing social media networks with NodeXL: Insights from a connected world.* Morgan Kaufmann.

Hartman, N. 2015. "VW needs massive marketing campaign to regain consumer trust – and survive." *The Concersation,* September 25. Accessed 02 February 2021. https://theconversation.com/vw-needs-massive-marketing-campaign-to-regain-consumer-trust-and-survive-48147

Ibrahim, N. F., X. Wang, and H. Bourne. 2017. "Exploring the effect of user engagement in online brand communities: Evidence from Twitter." *Computers in Human Behavior* 72: 321-338.

Ibrahim, N. F., and X. Wang. 2019. "Decoding the sentiment dynamics of online retailing customers: Time series analysis of social media." *Computers in Human Behavior* 96: 32-45.

James, E. H., L. P. Wooten, and K. Dushek. 2011. "Crisis management: Informing a new leadership research agenda." *The Academy of Management Annals* 5(1): 455-493.

Ji, X., S. A. Chun, Z. Wei, and J. Geller. 2015. "Twitter sentiment classification for measuring public health concerns." *Social Network Analysis and Mining* 5(1): 1-25.

Jin, S. A. A., and J. Phua. 2014. "Following Celebrities' Tweets About Brands: The Impact of Twitter-Based Electronic Word-of-Mouth on Consumers' Source Credibility Perception, Buying Intention, and Social Identification With Celebrities." *Journal of Advertising* 43(2): 181-195.

Ki, E. J., and E. Nekmat. 2014. "Situational crisis communication and interactivity: Usage and effectiveness of Facebook for crisis management by Fortune 500 companies." *Computers in Human Behavior* 35: 140-147.

Kim, E. H. J., Y. K. Jeong, Y. Kim, K. Y. Kang, and M. Song. 2016. "Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news." *Journal of Information Science* 42(6): 763-781.

Kim, Y., R. Dwivedi, J. Zhang, and S. R. Jeong. 2016. "Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5." *Online Information Review* 40(1): 42-61.

Kogan, M., L. Palen, and K. M. Anderson. 2015. "Think local, retweet global: Retweeting by the geographically-vulnerable during hurricane sandy." CSCW 2015: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing: 981-993.

Lachlan, K. A., P. R. Spence, X. Lin, K. Najarian, and M. Del Greco. 2016. "Social media and crisis management: CERC, search strategies, and Twitter content." *Computers in Human Behavior* 54: 647-652.

Lane, E. L. 2016. "Volkswagen and the high-tech greenwash." *European Journal of Risk Regulation* 7(1): 32-34.

Liau, B. Y., and P. P. Tan. 2014. "Gaining customer knowledge in low cost airlines through text mining." *Industrial Management & Data Systems* 114(9): 1344-1359.

Liu, X., S. Wang, K. Yao, and R. Sun. 2021. "Opportunistic behaviour in supply chain finance: a social media perspective on the 'Noah event'." *Enterprise Information Systems* 15(10): 1607-1634.

Ma, J., Y. K. Tse, X. Wang, and M. Zhang. 2019. "Examining customer perception and behaviour through social media research–An empirical study of the United Airlines overbooking crisis." *Transportation Research Part E: Logistics and Transportation Review* 127: 192-205.

Ma, J., Y. K. Tse, Y. Sato, M. Zhang, and Z. Lu. 2021. "Exploring the social broadcasting crisis communication: insights from the mars recall scandal." *Enterprise Information Systems* 15(3): 1-24.

Martinez, C. A., and J. D. Bowen. 2013. "The Ethical Challenges of the UN's Clean Development Mechanism." *Journal of Business Ethics* 117(4): 807-821.

Mathiesen, K., and A. Neslen. 2015. "VW scandal caused nearly 1m tonnes of extra pollution, analysis shows." *The Guardian*, Septermber 23. Accessed 11 February 2021. https://www.theguardian.com/business/2015/sep/22/vw-scandal-caused-nearly-1m-tonnes-of-extra-pollution-analysis-shows

McLeod, M. S., G. T. Payne, and R. E. Evert. 2016. "Organizational ethics research: A systematic review of methods and analytical techniques." *Journal of Business Ethics* 134(3): 429-443.

Moriarty, J. 2016. "Business Ethics." *The Stanford Encyclopedia of Philosophy,* November 17. Accessed 02 March 2021. https://plato.stanford.edu/entries/ethics-business/

Ngai, E. W., S. S. Tao, and K. K. Moon. 2015. "Social media research: Theories, constructs, and conceptual frameworks." *International Journal of Information Management* 35(1): 33-44.

Oh, O., M. Agrawal, and H. R. Rao. 2013. "Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets during Social Crises." *Mis Quarterly* 37(2): 407-426.

Omilion-Hodges, L. M., and K. L. McClain. 2016. "University use of social media and the crisis lifecycle: Organizational messages, first information responders' reactions, reframed messages and dissemination patterns." *Computers in Human Behavior* 54: 630-638.

Panagiotopoulos, P., J. Barnett, A. Z. Bigdeli, and S. Sams. 2016. "Social media in emergency management: Twitter as a tool for communicating risks to the public." *Technological Forecasting and Social Change* 111: 86-96.

Payne, G. T., C. B. Moore, R. G. Bell, and M. A. Zachary. 2013. "Signaling organizational virtue: an examination of virtue rhetoric, country-level corruption, and performance of foreign IPOs from emerging and developed economies." *Strategic Entrepreneurship Journal* 7(3): 230-251.

Pew Research Center. 2019. "Social Networking Fact Sheet." *Pew Research Center,* April 07. Accessed 09 March 2021. http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

Poulis, A., I. Rizomyliotis, and K. Konstantoulaki. 2019. "Do firms still need to be social? Firm generated content in social media." *Information Technology & People* 32(2): 387-404.

Quarshie, A. M., A. Salmi, and R. Leuschner. 2016. "Sustainability and corporate social responsibility in supply chains: The state of research in supply chain management and business ethics journals." *Journal of Purchasing and Supply Management* 22(2): 82-97.

Ramanathan, U., N. Subramanian, and G. Parrott. 2017. "Role of social media in retail network operations and marketing to enhance customer satisfaction." *International Journal of Operations & Production Management* 37 (1): 105-123.

Ranco, G., D. Aleksovski, G. Caldarelli, M. Grcar, and I. Mozetic. 2015. "The Effects of Twitter Sentiment on Stock Price Returns." *PLoS One* 10(9): e0138441.

Renn, O., 1991. "Risk communication and the social amplification of risk." In: Kasperson, R. E., P. J. M. Stallen. (Eds.), Communicating Risks to the Public. Kluwer Academic Publishers: 287-324.

Reuters. 2016. "Factbox: Timeline of VW auto emissions scandal." *The Reuters,* March 03. Accessed 29 November 2020. https://www.reuters.com/article/us-volkswagen-emissions-factbox-idUSKCN0W518W

Salama, M., H. A. Kader, and A. Abdelwahab. 2021. "An analytic framework for enhancing the performance of big heterogeneous data analysis." *International Journal of Engineering Business Management* 13: 1-11.

Schuler, D., A. Rasche, D. Etzion, and L. Newton. 2017. "Corporate Sustainability Management and Environmental Ethics." *Business Ethics Quarterly* 27(02): 213-237.

Shao, R., D. E. Rupp, D. P. Skarlicki, and K. S. Jones. 2013. "Employee justice across cultures: A meta-analytic review." *Journal of Management* 39(1): 263-301.

Sheng, J., J. Amankwah-Amoah, and X. Wang. 2017. "A multidisciplinary perspective of big data in management research." *International Journal of Production Economics* 191: 97-112.

Shiau, W. L., Y. K. Dwivedi, and H. S. Yang. 2017. "Co-citation and cluster analyses of extant literature on social networks." *International Journal of Information Management* 37(5): 390-399.

Shirky, C. 2011. "The Political Power of Social Media Technology, the Public Sphere, and Political Change." *Foreign Affairs* 90(1): 28-41.

Siano, A., A. Vollero, F. Conte, and S. Amabile. 2017. ""More than words": Expanding the taxonomy of greenwashing after the Volkswagen scandal." *Journal of Business Research* 7: 27-37.

Soltani, B. 2014. “The anatomy of corporate fraud: A comparative analysis of high profile American and European corporate scandals.” *Journal of Business Ethics* 120(2): 251-274.

Spence, P. R., K. A. Lachlan, and A. M. Rainear. 2016. “Social media and crisis research: Data collection and directions.” *Computers in Human Behavior* 54: 667-672.

Steinmeier, M. 2016. “Fraud in sustainability departments? An exploratory study.” *Journal of Business Ethics* 138(3): 477-492.

Stieglitz, S., Mirbabaie, M., Ross, B. and Neuberger, C. 2018. “Social media analytics–Challenges in topic discovery, data collection, and data preparation.” *International Journal of Information Management* 39:156-168.

Tajudeen, F. P., N. I. Jaafar, and S. Ainin. 2018. “Understanding the impact of social media usage among organizations.” *Information & Management* 55(3): 308-321.

Terpstra, T., R. Stronkman, A. de Vries, and G. L. Paradies. 2012. “Towards a Realtime Twitter Analysis during Crises for Operational Crisis Management.” ISCRAM 2012: In Proceedings of the 9th International ISCRAM Conference. Vancouver, Canada.

Tess, P. A. 2013. “The role of social media in higher education classes (real and virtual) - A literature review.” *Computers in Human Behavior* 29(5): A60-A68.

Thelwall, M., K. Buckley, and G. Paltoglou. 2011. “Sentiment in Twitter events.” *Journal of the American Society for Information Science and Technology* 62(2): 406-418.

Thelwall, M., K. Buckley, and G. Paltoglou. 2012. “Sentiment strength detection for the social web.” *Journal of the American Society for Information Science and Technology* 63(1): 163-173.

Thomaz, G. M., A. A. Biz, E. M. Bettoni, L. Mendes-Filho, and D. Buhalis. 2017. “Content mining framework in social media: A FIFA world cup 2014 case analysis.” *Information & Management* 54(6): 786-801.

Topham, G. 2015. “Volkswagen scandal – seven days that rocked the German carmaker.” *The Guardian,* September 25. Accessed 19 January 2021. https://www.theguardian.com/business/2015/sep/25/vw-emissions-scandal-seven-days

Trompeter, G. M., T. D. Carpenter, N. Desai, K. L. Jones, and R. A. Riley. 2013. "A synthesis of fraud-related research." *Auditing: A Journal of Practice & Theory* 32: 287–321.

Tse, Y. K., M. Zhang, B. Doherty, P. Chappell, and P. Garnett. 2016. "Insight from the horsemeat scandal: Exploring the consumers' opinion of tweets toward Tesco." *Industrial Management & Data Systems* 116(6): 1178-1200.

Tse, Y. K., H. Loh, J. Ding, and M. Zhang. 2018. "An investigation of social media data during a product recall scandal." *Enterprise Information Systems* 12(6): 733-751.

Vergeer, M., and P. H. Franses. 2016. "Live audience responses to live televised election debates: Time series analysis of issue salience and party salience on audience behavior." *Information, Communication & Society* 19(10): 1390-1410.

Zhao, X., M. Zhan, and C. Jie. 2018. "Examining multiplicity and dynamics of publics' crisis narratives with large-scale Twitter data." *Public Relations Review* 44(4): 619-632.

Zhu, L., D. Anagondahalli, and A. Zhang. 2017. "Social media and culture in crisis communication: McDonald's and KFC crises management in China." *Public Relations Review* 43(3): 487-492.

Table 1. Word Frequency of Stage 1

	FREQUEN	%SHOW	%PROCES	%TOT
CAR	868	3.78%	1.77%	0.86%
EMISSION	774	3.37%	1.58%	0.76%
COMPANY	671	2.92%	1.37%	0.66%
HITLER	506	2.20%	1.03%	0.50%
ADOLF	479	2.09%	0.98%	0.47%
DARK	478	2.08%	0.98%	0.47%
FOUNDER	477	2.08%	0.97%	0.47%
LEGACY	476	2.07%	0.97%	0.47%
STAIN	475	2.07%	0.97%	0.47%
DAVESFOLEY	474	2.06%	0.97%	0.47%
GERMAN	438	1.91%	0.89%	0.43%
DIESEL	418	1.82%	0.85%	0.41%
AMP	354	1.54%	0.72%	0.35%
TDI	269	1.17%	0.55%	0.27%
INDUSTRY	241	1.05%	0.49%	0.24%
YEAR	200	0.87%	0.41%	0.20%
AUTO	179	0.78%	0.37%	0.18%
DATA	177	0.77%	0.36%	0.17%
CEO	172	0.75%	0.35%	0.17%
SHOWS	168	0.73%	0.34%	0.17%
CLIMATE	162	0.71%	0.33%	0.16%
NEWS	155	0.67%	0.32%	0.15%
ENGINEER	151	0.66%	0.31%	0.15%
PEOPLE	146	0.64%	0.30%	0.14%
SOFTWARE	146	0.64%	0.30%	0.14%
CHEATING	145	0.63%	0.30%	0.14%
TIME	143	0.62%	0.29%	0.14%
BUY	135	0.59%	0.28%	0.13%
GOOD	131	0.57%	0.27%	0.13%
TRUST	131	0.57%	0.27%	0.13%
EPA	125	0.54%	0.26%	0.12%
WINTERKORN	125	0.54%	0.26%	0.12%

Note: Some words, such as "Volkswagen", "scandal", "VWgate", "Volkswagenscandal", "UKVolkswagen", "VWscandal" and "VW" are excluded, as they are the search criteria of the dataset. "RT" is also not included as it is not relevant to the crisis.

Table 2. Word Frequency of Stage 2

	FREQUEN	%SHOW	%PROCES	%TOT
CAR	3767	6.36%	2.29%	1.15%
EMISSION	2712	4.58%	1.65%	0.83%
GERMAN	1822	3.08%	1.11%	0.56%
DIESEL	1296	2.19%	0.79%	0.40%
AMP	1132	1.91%	0.69%	0.35%
TDI	1059	1.79%	0.64%	0.32%
AUTO	915	1.54%	0.56%	0.28%
SOFTWARE	803	1.36%	0.49%	0.25%
AUDI	786	1.33%	0.48%	0.24%
MAKE	740	1.25%	0.45%	0.23%
LINK	760	1.28%	0.46%	0.23%
REMOVED	744	1.26%	0.45%	0.23%
CHEATING	733	1.24%	0.45%	0.22%
UKHELP	652	1.10%	0.40%	0.20%
STREET	632	1.07%	0.38%	0.19%
WALL	624	1.05%	0.38%	0.19%
TESTS	616	1.04%	0.37%	0.19%
JOURNAL	620	1.05%	0.38%	0.19%
OWNERS	604	1.02%	0.37%	0.18%
EU	511	0.86%	0.31%	0.16%
INDUSTRY	519	0.88%	0.32%	0.16%
MILLION	502	0.85%	0.31%	0.15%
DIESELGAT	488	0.82%	0.30%	0.15%
COMPANY	478	0.81%	0.29%	0.15%
NEWS	475	0.80%	0.29%	0.15%
HIT	478	0.81%	0.29%	0.15%
TIME	469	0.79%	0.29%	0.14%
GROUP	463	0.78%	0.28%	0.14%
HOPE	454	0.77%	0.28%	0.14%
UK	445	0.75%	0.27%	0.14%
GOLF	438	0.74%	0.27%	0.13%
MOVIE	429	0.72%	0.26%	0.13%

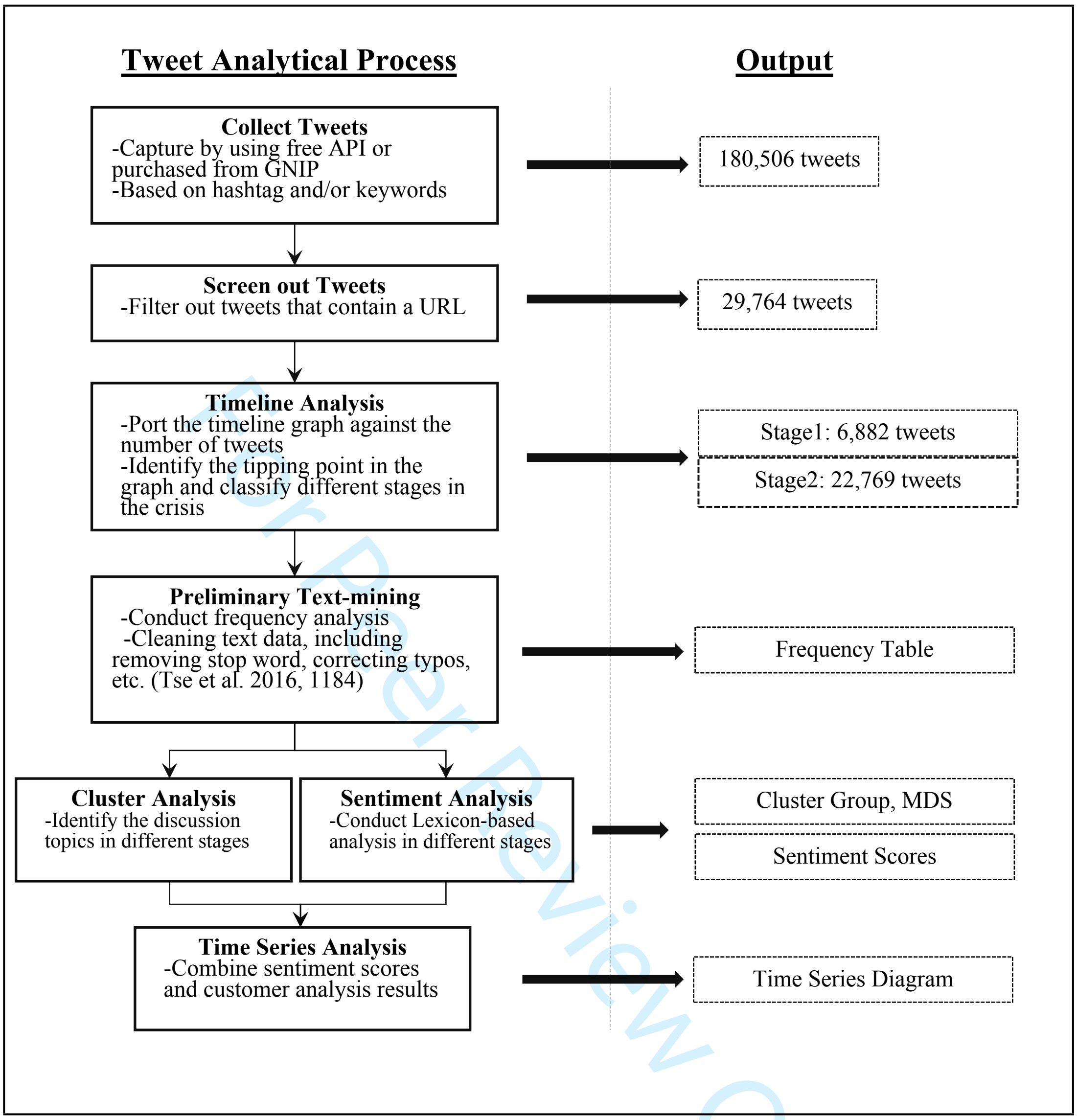

Figure 1. Tweet Analytical Framework

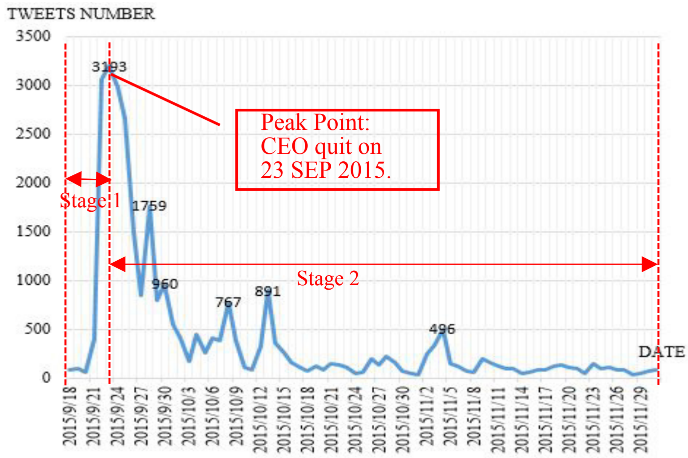

Figure 2. Time Trend of Volume of Tweets

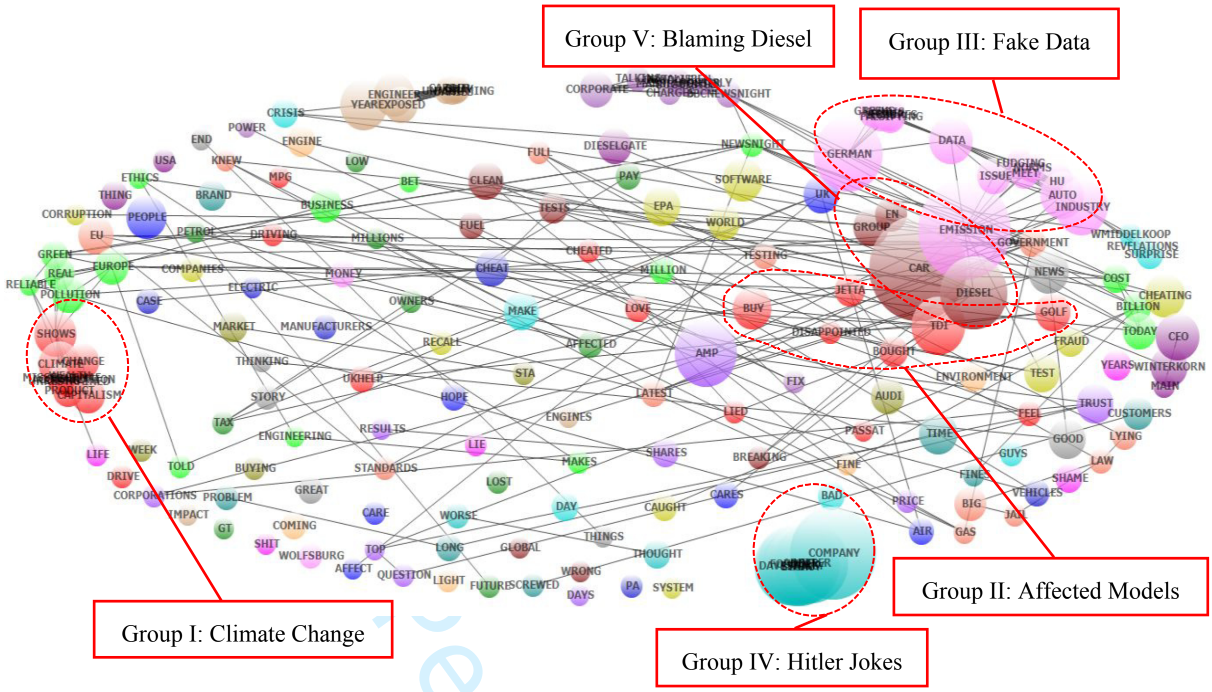

Figure 3. The Mapping of Stage 1

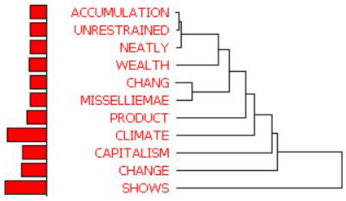

Figure 4. Dendrogram of Climate Change group

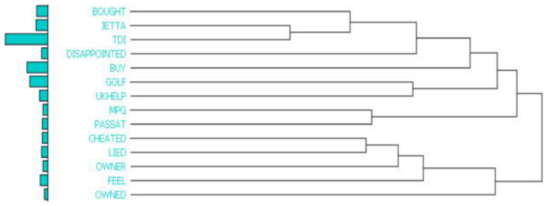

Figure 5. Dendrogram of Affected Models group

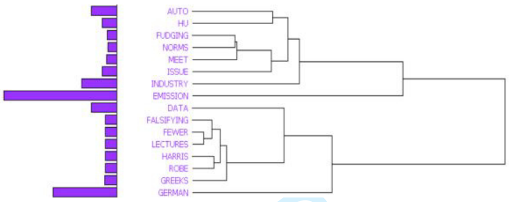

Figure 6. Dendrogram of Fake Data group

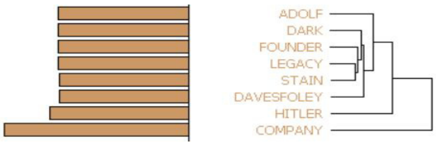

Figure 7. Dendrogram of Hitler Jokes group

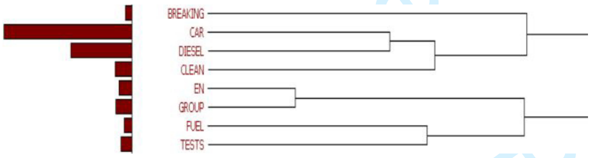

Figure 8. Dendrogram of Blaming Diesel group

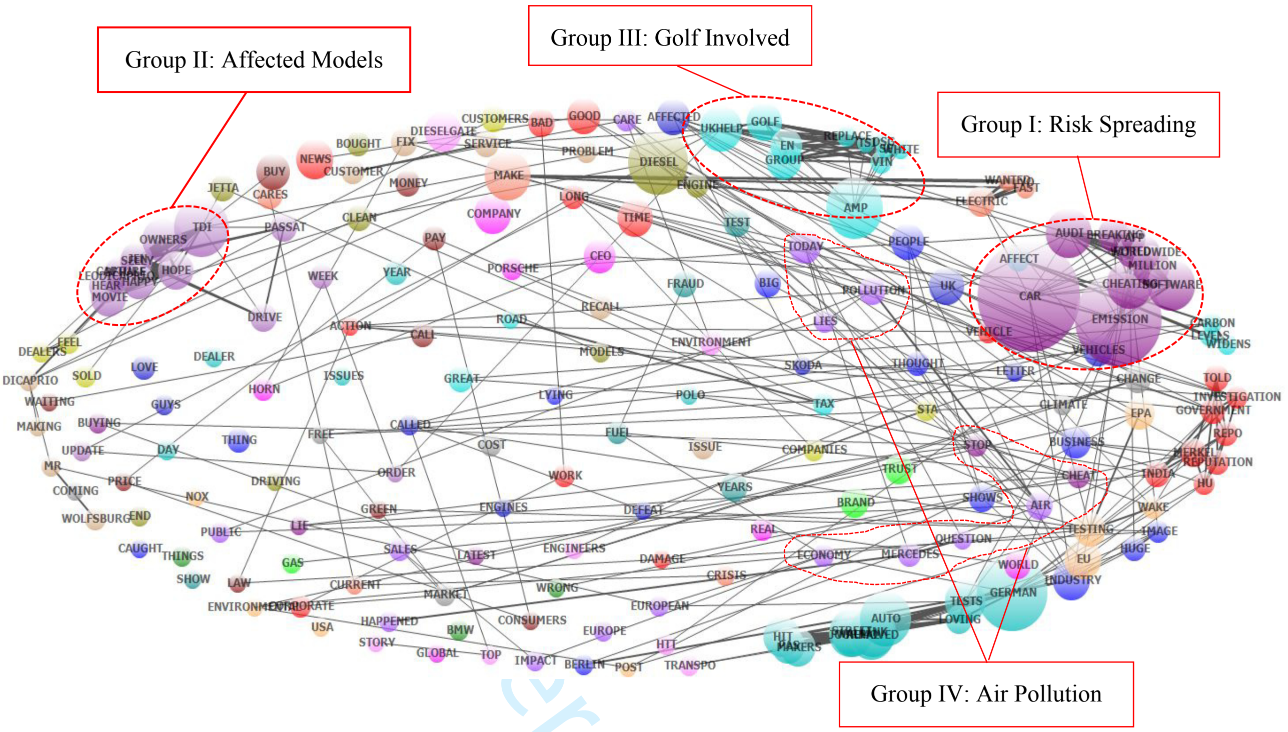

Figure 9. The Mapping of Stage 2

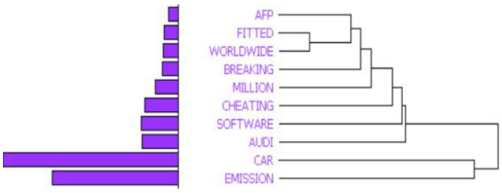

Figure 10. Dendrogram of Risk Spreading group

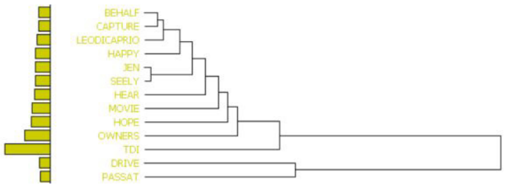

Figure 11. Dendrogram of Affected Models group

Figure 12. Dendrogram of Golf Involved group

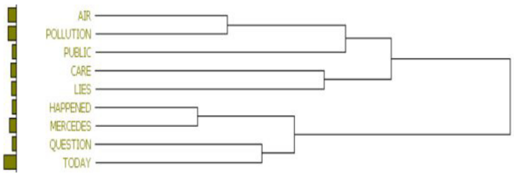

Figure 13. Dendrogram of Air Pollution group

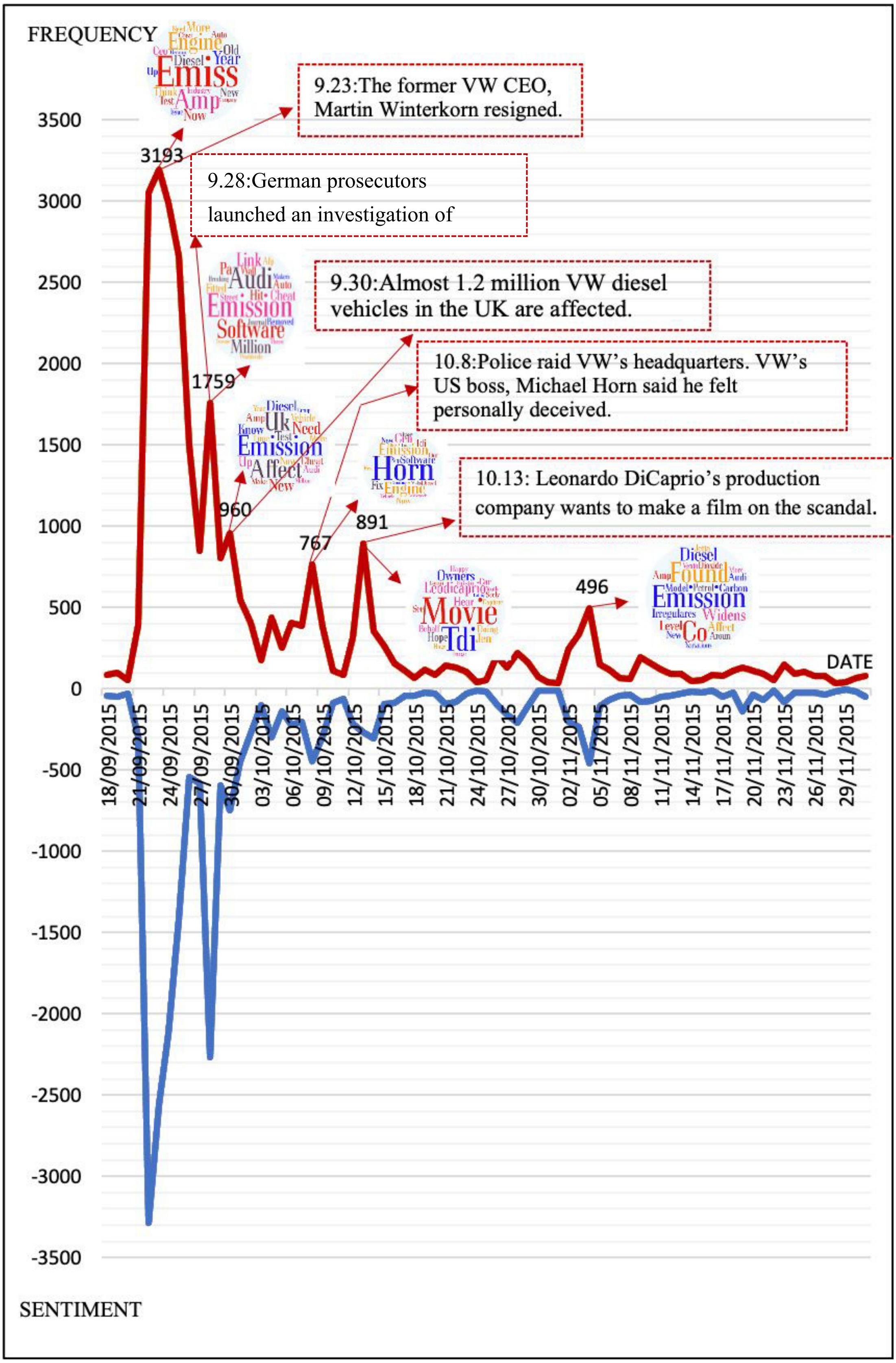

Figure 14. Time Series for Frequency and Sentiment

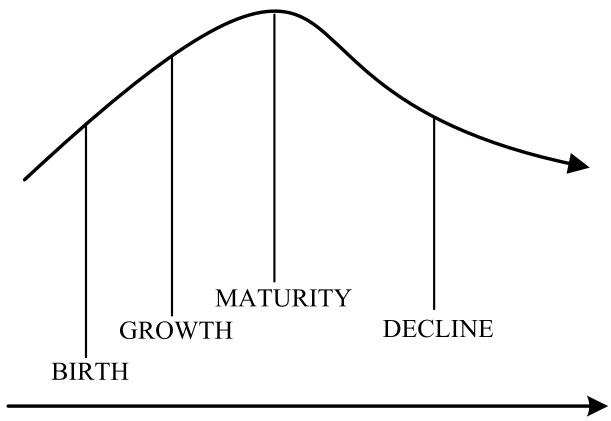

Figure 15. The Crisis Life Cycle; Adopted from González-Herrero and Pratt (1996)