Text Mining Applications to Brands' Social Media Messages Pre-, During and Post-Pandemic

Mustafa Cerit
Turk Alman Universitesi
Elif Karaosmanoglu
Istanbul Technical University

Cite as:
Cerit Mustafa, Karaosmanoglu Elif (2024), Text Mining Applications to Brands' Social Media Messages Pre-, During and Post-Pandemic. *Proceedings of the European Marketing Academy*, (122611)

Paper from EMAC Regional Conference, Lisbon, Portugal, September 25-27, 2024

Text Mining Applications to Brands' Social Media Messages Pre-, During and Post-Pandemic

Abstract

By applying text-mining techniques to tweets of companies in Turkey, we aim to examine and to contrast the alterations in the businesses' communication from before to after the Covid-19 pandemic. By using Structural Topic Model, we found that online transactions and communication campaigns emerged as prominent marketing-related topics. We broaden our investigation by analyzing the emotional patterns of the identified themes and found notable variations in emotional patterns across communication campaigns, depending on the time frame. However, we did not see any major variances in the discourses linked to online transactions. Furthermore, while trust, anticipation, and joy are consistently desired emotions in brand messages regardless of the situation, fear, sadness, anger, disgust, and surprise vary depending on the time of the pandemic.

Keywords: *marketing communication, text-mining, Covid-19 pandemic*

1. Introduction

During the COVID-19 pandemic, automated text mining techniques are widely used in marketing literature to gain insights into marketing communication (Farmaki et al., 2022; Mangiò et al., 2021; Pezoa-Fuentes, 2023) or consumer behavior. For instance, one study found the ethical and social tones in social media messages of brands during the global pandemic by applying certain text mining to these messages (Mangiò et al., 2021). Other research showed that automated text mining may discover pandemic-related information in brand social media posts (Hesse et al., 2021).

Despite the many studies on user-generated content on social media (Naurin & Michel, 2023; Yun Kyung & Jisu, 2022; Yousefi et al., 2020), to our knowledge, there is no comparative evidence on the use of text-mining techniques to analyze brand discourses before, during, and after the pandemic. This research is important because the pandemic is a rare type of crisis that can affect many companies, and because such crises are likely to occur in the future. Thus, studying pandemic-focused social media communications may help marketers prepare for future crises (Hesse et al., 2021). Scheiwiller and Zizka (2021) underline the importance of evaluating communications before, during, and after the pandemic to see how much the content differed from other issues. To contribute to this need, this paper investigates brand messages posted on social media before, during, and after the pandemic by applying certain text mining methods to brands'messages posted on Twitter over three years from 2019 to 2022.

In order to accomplish this, we initially retrieved our data from Twitter by utilizing the R programming language. Following the completion of data preparation operations, we used structural topic model (STM) to investigate the major topics discussed over the years in question. Then, we conducted an in-depth research of business-related subjects by utilizing sentiment analysis to identify the underlying emotional patterns. Our findings provide insight into how marketing messages might be modified based on the specific conditions of societal crises. Furthermore, we have delineated the prominent emotions that may arise within brand communications during such crises.

The structure of this document is as follows. Initially, we will provide an explanation of the data collection process and further expand on the procedures involved in data preparation. Afterwards, we go on to emotional analysis following the investigation of STM. In conclusion, we will discuss brand communications during times of crisis, which aligns with the findings of our research.

2. Data Collection

Twitter is selected as the source of social media data because it serves as an integrated marketing communications platform for engaging and informing both current and potential consumers (Liu, 2020). During the pandemic, organizations employed Twitter as a means to disseminate information to the public regarding developments relating to the disaster (Dhar & Bose, 2022). We utilized the Twitter API to acquire tweets for the purpose of data analysis. The R programming language is utilized to gather and analyze the accumulated tweets.

The sampling source for identifying the brands to analyze consisted of three different institutions' lists, namely, Brand Finance's "Türkiye 100 2022," Capital's "The Most Valuable Brands of Türkiye 2022," and Ipsos' "Lovemarks of Türkiye 2021." While certain brands, such as "Arçelik" and "Türkiye İş Bankası," are included multiple times, others like "Opet" and "Aytemiz" are only listed once. We extracted certain accounts, such as the inactive ones, football clubs, and the Istanbul Stock Exchange, whose contents were not suitable for the research goal. We obtained 85 distinct brands at the end. The final dataset comprised 22,803 tweets and associated metrics. Only tweets in Turkish were utilized for this purpose.

3. Data Preprocessing

Following the approval to access the Twitter API, we employed the R software together with the *twitteR* (Gentry, 2022) and *rtweet* (Kearney et al., 2024) packages to gather the tweets from the designated brands. After gathering the tweets, we carried out the text preprocessing procedures described by Berger et al. (2020), which involved eliminating punctuation marks, numbers, and stop words. Additionally, we also used Regular Expressions (*regex*) to remove certain words by targeting specific patterns in order to refine the data set.

After finishing the process of data preprocessing, the full dataset was partitioned into three distinct time periods. The official declaration of the pandemic in Türkiye took place on March 11, 2020. We utilized this date as our point of reference. We defined the time span from March 11, 2019 to March 11, 2020 as "prior to the pandemic", while we categorized the period from March 11, 2020 to March 11, 2021 as "during the pandemic". We labeled the period from March 11, 2021 to March 11, 2022 as "after the pandemic". Prior to the pandemic, the brands in the sample accumulated a total of 5377 Tweets, however during the pandemic, the brands tweeted a total of 7406 times. The total number of tweets following the pandemic is 10020.

4. Structural Topic Model (STM)

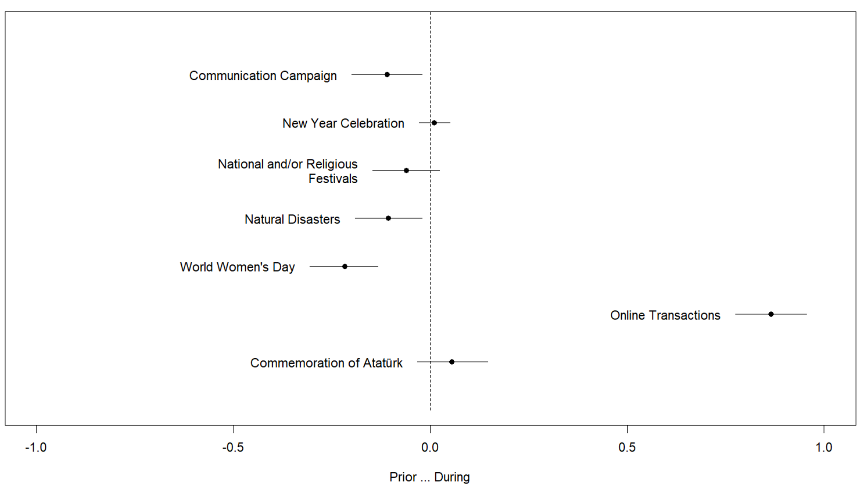
Topic modeling approaches are commonly employed for automatically discovering hidden subjects in large amounts of text by establishing associations between terms in the text (Berger et al., 2020). The Structural Topic Model (STM) differs from existing topic models like Latent Dirichlet Allocation (LDA) (Blei et al., 2003) and Correlated Topic Model (Blei & Lafferty, 2007) by integrating the metadata of the entire dataset into its analytic process. This allows researchers to create connections between the metadata and the data.

Before initiating STM analysis, we gathered tweets that were posted on the same day across a span of three years and categorized them into certain time periods, namely pre-COVID, during, and post-COVID. Afterwards, we created two subsets: one consisting of tweets that were published before and during the epidemic, and the other containing tweets that were released during and after the pandemic in order to carry out binary comparisons within each subset. Our primary objective is to examine the changes in brands' marketing discourse on social media during the pandemic compared to the years before and after the pandemic. To achieve this, we conducted a comparative analysis of these two time periods. To implement STM model, we determined the specific number of subjects to be identified by STM upon discussion by evaluating the identified topics from the number of 5 to 25 in terms of interpretability of each topic. Ultimately, we concluded that a set of 7 topics effectively captures all the themes included in tweets shared prior to and during the epidemic, whereas a set of 12 topics is the most suitable for tweets shared during and after the pandemic.

4.1. Before and during the pandemic

Based on the information provided earlier, we have determined that having 7 topics is the most suitable amount to encompass all of the distinct subjects that have arisen both before and during the pandemic. In particular, the topics that we named are “communication campaigns”, “the New Year celebration”, “national and/or religious festivals”, “the earthquake”, “International Women's Day”, “online transactions during the pandemic”, and “Commemoration of Mustafa Kemal Atatürk”, the founder of the Republic of Turkey. Figure 1 demonstrates the correlation between the specified subjects and the time periods under investigation. The majority of social media messages pertaining to the global pandemic contained phrases that referred to online transactions. We observed that businesses integrated messages to elucidate to consumers the ease with which they could conduct their transactions using mobile applications or websites during the pandemic. Conversely, the significance of communication campaigns diminished during the pandemic in comparison to the pre-pandemic era. In addition, brands have consistently shown sensitivity towards social events, such as natural disasters, as exemplified by the topic "Earthquake," and special days, such as "International Women's Day," "New Year," "Greeting Message for the Festival," and "Commemoration of Atatürk."

Figure 1 Metadata and Topic Relations prior and during COVID-19

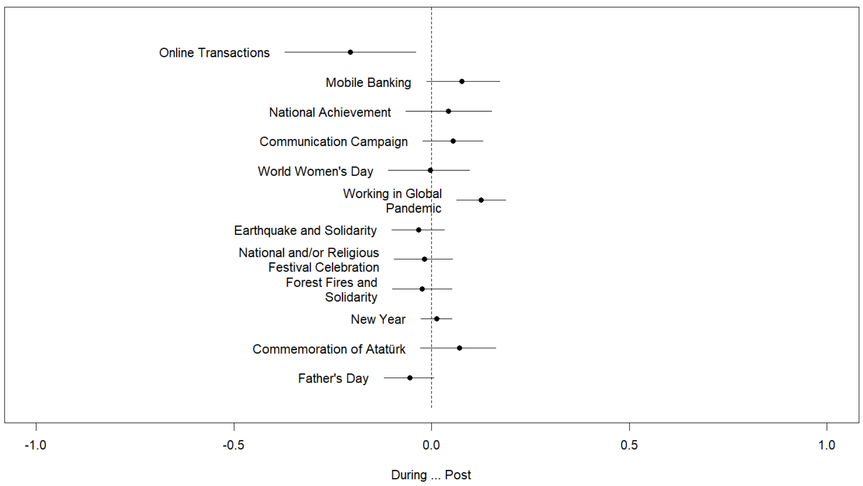


4.2. During and after the pandemic

After conducting multiple trials, we have determined that 12 themes is the ideal amount to encompass all the distinct subjects discussed in the brand's Twitter accounts during and after the pandemic. The topics that we defined are “online transactions”, “mobile banking”, “national achievement in Olympic games”, “seasonal communication campaign”, “International Women's Day”, “working conditions during the pandemic, “earthquakes”, “Forest Fires”, “National and/or Religious Festival Celebration”, “the New Year celebration”, “Commemoration of Mustafa

Kemal Atatürk", "Father's Day". Figure 2 presents that Turkish banks were increasingly utilizing Twitter to advertise the convenience of conducting online banking operations. The value of these signals has been increasingly noticeable, particularly following the global pandemic, while the impact of online transactions has decreased. In addition, the significance of communication campaigns became significant again after the pandemic. Furthermore, organizations often addressed the working environment during the pandemic and conveyed appreciation for the employees who continued to work despite the health risks.

Figure 2 Metadata and Topic Relations during and after COVID-19



5. Emotion Patterns' Analysis

In the following stage, we examined the general emotional content associated with specific topics and discovered the fluctuations in these emotions across the years in question. To achieve this, we utilized the Turkish edition of the NRC Word-Emotion Association Lexicon (Mohammad & Turmey, 2013). This lexicon consists of 13,872 words that are classified into two sentiments and eight emotions: positive, negative, trust, surprise, fear, anticipation, joy, anger, disgust, and sadness. According to the initial investigation, it has been constant irrespective of the context that positive tones have overwhelmingly prevailed in all emotion-based tweets and that negative sentiments were comparatively infrequent within each time period. Therefore, we have derived positive and negative sentiments from the vocabulary in order to analyze the alterations in the remaining emotions. In order to analyze brand messaging pertaining to marketing activity during the pandemic, we selected the common themes which are seen in both time periods. These are "communication campaigns" and "online transactions during the pandemic". In addition, we considered "mobile banking" as kind of digital transactions as well, therefore we took this topic also as a common theme to proceed with the further analyses.

Prior to visually evaluating the relative changes in the emotions associated with the above topics, we performed a Mann-Whitney U test for each pair of time periods to determine if the changes were statistically significant. To implement the test, we calculated the total number of terms that represent specific emotions. This allowed us to obtain eight distinct values for each of the two time periods, as shown in Table 1.

Table 1 Data Set used for the Mann-Whitney U test

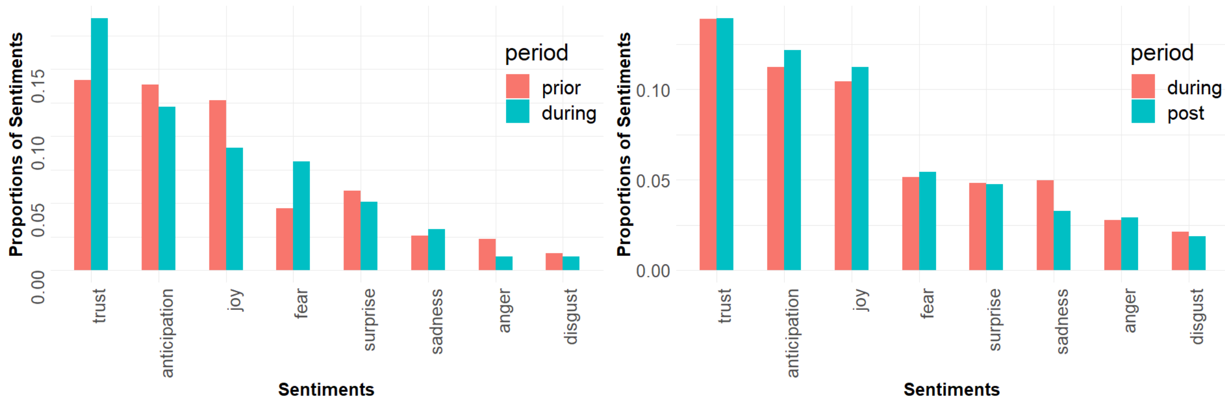
Emotion	Sum of the Terms' Count referring to the Emotion	Period	Sum of the Terms' Count referring to the Emotion	Period	Sum of the Terms' Count referring to the Emotion	Period
	Communication Campaigns		Seasonal Communication Campaigns		Online Transactions during the pandemic + Mobile Banking during the pandemic	
Anger	122	prior	35	during	863	during
anticipation	732	prior	142	during	3397	during
Disgust	67	prior	27	during	459	during
Fear	245	prior	65	during	1481	during
Joy	670	prior	132	during	3198	during
Sadness	137	prior	63	during	883	during
Surprise	312	prior	61	during	1493	during
Trust	750	prior	176	during	3828	during
Anger	2	during	225	after	341	after
anticipation	24	during	938	after	1586	after
Disgust	2	during	145	after	167	after
Fear	16	during	418	after	622	after
Joy	18	during	867	after	1453	after
Sadness	6	during	253	after	333	after
Surprise	10	during	365	after	685	after
Trust	37	during	1073	after	1787	after

Table 1 indicates for instance that there are 122 words related to anger in the tweets posted before the pandemic. These tweets fall under the topic of "communication campaigns". The number of words expressing trust in the tweets related to topics of "Online Transactions during the pandemic" and "Mobile Banking during the pandemic" is 3828.

5.1. Communication campaigns

As anticipated, corporations have persistently launched communication initiatives during the designated time frame. Figure 3 demonstrates the comparative significance of emotions in brand messages released over the years under investigation. The utilization of emotions in the communications varies, as depicted in the figures, and these variances are also statistically significant (before and during the pandemic: U (N_{prior} = 8, N_{during} = 8) =.500, z = -3.313, p < 0.001; during and after the pandemic: U (N_{during} = 8, N_{post} = 8) = 1.000, z = -3.256, p < 0.001.)

Figure 3 Sentiments in Marketing Messages



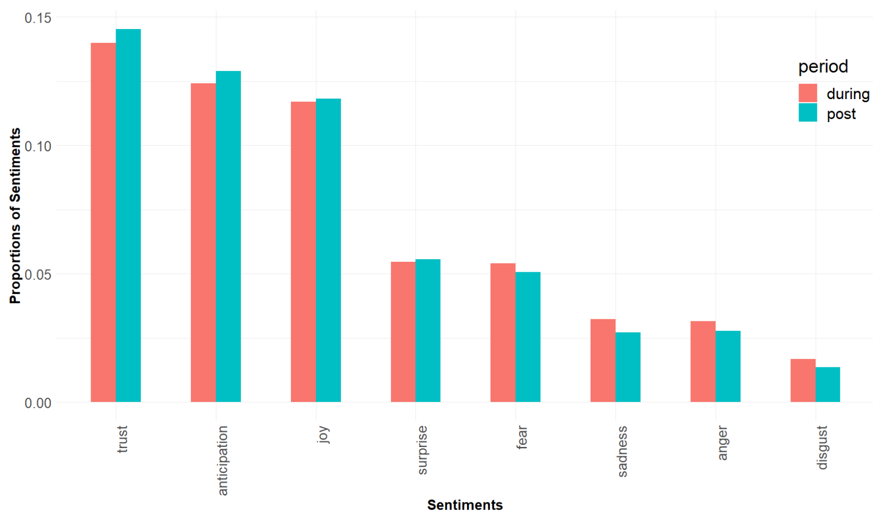
The left side of Figure 3 shows that the most prevalent feelings found were trust, anticipation, and joy, regardless of the year being studied. Further, we observed that brand communications have placed a higher emphasis on phrases associated with trust during the pandemic. Similarly, there was a higher occurrence of terms associated to fear compared to before the pandemic. On the other hand, the use of words associated with anticipation, joy, and surprise declined throughout the pandemic. Although terms associated with sadness, anger, and disgust were not as frequently used as terms related to other emotions, we noticed a small rise in the importance of sadness in brand messages, while the usage of anger and disgust-related terms decreased.

Upon analyzing the two-year period that included the pandemic and its aftermath, we have identified a consistent pattern about the prevailing emotions noticed during that time. The right side of Figure 3 illustrates that the companies used joy, anticipation, and trust as the main emotions in their communications over both time periods. The level of anticipation and joy experienced a rise from one year to the next. In the same way, the proportions of phrases associated to fear increased even after the pandemic, suggesting that the impact of the global pandemic was still present in the brand's communications even after the pandemic. Furthermore, there has been a significant decrease in the prominence of terms associated with sadness from the period during the pandemic to the period after the pandemic. Although there is no variation in the relative weights of trust-related variables when comparing different time periods, trust consistently remains the dominant feeling in both times, as found in previously analyzed timeframes.

5.2. Online transactions and mobile banking

Although it is possible to observe changes in emotions mentioned in different time periods, as shown in Figure 4, the results of the Mann-Whitney U test indicate that there are no statistically significant differences in the emotions expressed in tweets posted during and after the pandemic (U (N_{during} = 8, N_{post} = 8) =15.ooo, z = -1.785, p = 0.083).

Figure 4 Sentiments of Digitalization-Related Messages



We found out that the tweets that we labeled "online transactions" in the timeframe prior to and during the pandemic were posted solely during the pandemic. Therefore we were not able to compare the emotive patterns of this topic within the first pair of time period. On the other hand, it was possible to compare the emotive patterns of topics "online transactions" and "mobile banking" belonging to the period from during to after the pandemic because the tweets were posted either during or after the pandemic. Figure 4 offers preliminary evidence supporting the assertion that brands consistently incorporated trust, anticipation, and joy as the main emotions in their Twitter posts, regardless of changes in context. Following the pandemic, brand messages have increasingly emphasized positive feelings. This can be related to the reduced proportionate utilization of fear, sadness, anger, and disgust, in comparison to the heightened proportionate usage of trust, anticipation, joy, and surprise.

6. Conclusion

Our findings indicate a notable difference in emotion patterns between the brand messages across the years in question. However, there was no similar difference in the topics related to digitalization when comparing the years before and after the pandemic. This finding indicates that corporations have modified their communication strategies for marketing messaging in response to the circumstances arising from the pandemic, while maintaining a rather consistent approach to their discussions on digitalization.

Our findings align with previous research, which shows that brands primarily use emotional patterns of joy, anticipation, and trust in their social media messaging during times of disruptive events. Moreover, a prevailing trend of positive attitude is regularly observed on all platforms (Pezoa-Fuentes et al., 2023). Consistent with recent research (Dhar & Bose, 2022), our study confirms that brands in Turkey place importance on maintaining continuity during the pandemic. They effectively communicate this idea through their business-focused social media posts. Furthermore, it has been demonstrated in these posts that in times of crisis, companies make an effort to actively involve the community and convey a sense of optimism (Arora et al., 2022). Similar to how trust becomes the most important factor influencing human behavior, especially during the pandemic (Arora et al., 2022), we too notice this dominant emotional pattern in the commercial messages we study. Mathayomchan et al. (2023) found that when the impact of the pandemic decreases, public discussions tend to contain more optimistic content, leading to a shift from negative to positive speech. The brand messages that were investigated during our analysis exhibit same trends.

Fear and trust are the main factors that shape public discourse during the pandemic. Our research also shows that companies use specific language, particularly related to trust, throughout the pandemic. However, the use of fear-related phrases increases significantly during times of crisis. According to the study conducted by Li and McCrary (2022), customers have generally experienced anticipation as a very favorable emotion during the pandemic, whereas surprise has been seen less favorably. The findings align with these results as well. Specifically, the use of surprise has decreased during the pandemic, while anticipation is consistently regarded as one of the most prevalent emotions conveyed in advertising communications. In addition to typical business communications, our study incorporates the findings of prior research (Farmaki et al., 2022) to demonstrate that the brands we examined specifically express their support for the victims of the pandemic and natural disasters. Moreover, the marketing messages frequently address the employee well-being, which aligns with the findings of other studies (Scheiwiller & Zizka, 2021).

The scarcity of extensive research on brand content on social media has left marketing professionals seeking direction on creating appropriate social media content (Pezzuti et al., 2021). Furthermore, this requirement becomes increasingly crucial during periods of crises. Research focused on company-generated material on social media has the potential to contribute to an expanding body of literature (Srinivasan et al., 2022). Our research intends to contribute to the need for understanding the impact of the pandemic on domestic and international brands operating in Turkey by using automated text mining tools on a large dataset. There is a shortage of comparative research in the literature that examine the periods before, during, and after the pandemic, while there are some studies available that explicitly focus on the pandemic period (Hesse et al., 2021; Farmaki et al., 2022; Mangiò et al., 2021). Our research also addresses this gap by applying its methods to three unique historical periods.

7. References

Arora, S., Debesay, J., & Eslen-Ziya, H. (2022). Persuasive narrative during the COVID-19 pandemic: Norwegian Prime Minister Erna Solberg's posts on Facebook. *Humanit Soc Sci Commun, 9*(35).

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, *84*(1), 1-25.

Blei, D. M., & Lafferty, J. D. (2007). A Correlated Topic Model of Science. *The Annals of Applied Statistics, 1*(1), 17-35.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 993-1022.

Dhar, S., & Bose, I. (2022). Victim crisis communication strategy on digital media: A study of the COVID-19 pandemic. *Decision Support Systems, 161*(October).

Farmaki, A., Hadjielias, E., Olya, H., Taheri, B., & Drotarova, H. M. (2022). CSR communication and international marketing: Insights from the COVID-19 pandemic. *Internaitonal Marketing Review*.

Gentry, J. (2022, October 14). Package "twitteR".

Hesse, A., Niederle, F., Schön, L., & Stautz, D. (2021). Consumer responses to brand communications involving COVID-19. *Journal of Marketing Management, 37*(17-18), 1783-1814.

Kearney, M. W., Wickham, H., Heiss, A., Briatte, F., & Sidi, J. (2024, February 24). Package "rtweet".

Li, J., & McCrary, R. (2022). Consumer communications and current events: a cross-cultural study of the change in consumer response to company social media posts due to the COVID-19 pandemic. *Journal of Marketing Analytics, 10*, 173-183.

Liu, X. (2020). Analysing the impact of user-generated content on B2B Firms' stock performance: Big data analysis with machine learning methods. *Industrial Marketing Management, 86*, 30-39.

Mangiò, F., Pedeliento, G., & Andreini, D. (2021). Branding Rhetoric in Times of a Global Pandemic: A Text-Mining Analysis. *Journal of Advertising, 50*(3), 240-252.

Mathayomchan, B., Taecharungroj, V., & Wattanacharoensil, W. (2023). Evolution of COVID-19 tweets about Southeast Asian Countries: topic modelling and sentiment analyses. *Place Brand and Public Dipl, 19*, 317-334.

Mohammad, S. M., & Turmey, P. D. (2013). Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*.

Naurin, A. S., & Michel, L. (2023). Extracting marketing information from product reviews: a comparative study of latent semantic analysis and probabilistic latent semantic analysis. *Journal of Marketing Analytics, 11*, 662-676.

Pezoa-Fuentes, C., García-Rivera, D., & Matamoros-Rojas, S. (2023). Sentiment and Emotion on Twitter: The Case of the Global Consumer Electronics Industry. *J. Theor. Appl. Electron. Commer. Res., 18*, 765-776.

Pezzuti, T., Leonhardt, J. M., & Warren, C. (2021). Certainty in Language Increases Consumer Engagement on Social Media. *Journal of İnteractive Marketing, 53*, 32-46.

Scheiwiller, S., & Zizka, L. (2021). Strategic responses by European airlines to the Covid-19 pandemic: A soft landing or turbulent ride? *Journal of Air Transport Management, 95*, 101863.

Schivinski, B., & Dabrowski, D. (2015). The impact of brand communication on brand equity through Facebook. *Journal of Research in Interactive Marketing, 9*(1), 31-53.

Srinivasan, R., Kumar, J. A., & Kumar, N. (2022). To talk or not?: An analysis of firm-initiated social media communication's impact on firm value preservation during a massive disruption across multiple firms and industries. *Decision Sciences*.

Yousefi, D. Z., Nastaran, H., & Rouhani, S. (2020). User response to e-WOM in social networks: how to predict a content influence in Twitter. *International Journal of Internet Marketing and Advertising, 14*(1), 91-111.

Yun Kyung, O., & Jisu, Y. (2022). Asymmetric effect of feature level sentiment on product rating: an application of bigram natural language processing (NLP) analysis. *Internet Research, 32*(3), 1023-1040.