

Article

Content and Sentiment Analysis of The New York Times Coronavirus (2019-nCOV) Articles with Natural Language Processing (NLP) and Leximancer

Sezai Tunca * , Bulent Sezen and Yavuz Selim Balcioglu 

Management Information System Department, Faculty of Management, Gebze Technical University, 41400 Kocaeli, Turkey

* Correspondence: stunca@gtu.edu.tr

Abstract: The purpose of this study was to prove the use of content and sentiment analysis to understand public discourse on Nytimes.com around the coronavirus (2019-nCOV) pandemic. We examined the pandemic discourses in the article contents, news, expert opinions, and statements of official institutions with natural language processing methods. We analyzed how the mainstream media (Nytimes.com) sets the community agenda. As a method, the textual data for the research were collected with the Orange3 software text-mining tool via the Nytimes.com API, and content analysis was conducted with Leximancer software. The research data were divided into three categories (first, mid, and last) based on the date ranges determined during the pandemic. Using Leximancer concept maps tools, we explained concepts and their relationships by visualizing them to show pandemic discourse. We used VADER sentiment analysis to analyze the pandemic discourse. The results gave us the distance and proximity positions of themes related to Nytimes.com pandemic discourse, revealed according to their conceptual definitions. Additionally, we compared the performance of six machine learning algorithms on the task of text classification. Considering the findings, it is possible to conclude that in Nytimes.com (2019-nCOV) discourse, some concepts have changed on a regular basis while others have remained constant. The pandemic discourse focused on specific concepts that were seen to guide human behavior and presented content that may cause anxiety to readers of Nytimes.com. The results of the sentiment analysis supported these findings. Another result was that the findings showed us that the contents of the coronavirus (2019-nCOV) articles supported official policies. It can be concluded that regarding the coronavirus (2019-nCOV), which has caused profound societal changes and has results such as death, restrictions, and mask use, the discourse did not go beyond a total of 15 main themes and about 100 concepts. The content analysis of Nytimes.com reveals that it has behavioral effects, such as causing fear and anxiety in people. Considering the media dependency of society, this result is important. It can be said that the agenda-setting of society does not go beyond the traditional discourse due to the tendency of individuals to use newspapers and news websites to obtain information.

Keywords: coronavirus (2019-nCOV); mainstream media; content analysis; natural language processing; sentiment analysis



Citation: Tunca, S.; Sezen, B.; Balcioglu, Y.S. Content and Sentiment Analysis of The New York Times Coronavirus (2019-nCOV) Articles with Natural Language Processing (NLP) and Leximancer. *Electronics* **2023**, *12*, 1964. <https://doi.org/10.3390/electronics12091964>

Academic Editor: Ruifeng Xu

Received: 4 December 2022

Revised: 25 December 2022

Accepted: 13 January 2023

Published: 23 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The coronavirus (2019-nCOV) pandemic, which is thought to have started in December 2019, has affected the global population's daily life, lifestyle, and behavior as it spread across the world [1]. The pandemic has had such an impact that it became the number one problem on the agendas of all countries [2]. All newspapers, magazines, online and social media news, statements, and articles about the pandemic were used as sources of information. According to the reports on Forbes.com, pandemic-related news has been especially prevalent in the social media environment. Looking at 28 February 2020, it was

estimated that 6.7 million people talked about the coronavirus (2019-nCoV) in one day, and as a result of this widespread discussion, people became worried [3]. Following this, people have relied on perceived high-quality news media to learn about the coronavirus (2019-nCoV) [4]. The coronavirus (2019-nCoV) crisis has led to significant increases in the consumption of news in digital media since the beginning of 2020, and people have felt the need to be extremely informed [5]. In conjunction with this, news reading behaviors have rapidly increased since the onset of the pandemic [6]. Moreover, it has been determined that people who turn to alternative media for news also use the mainstream media [7]. Nytimes.com, which is a part of the mainstream media, has shared articles, news stories, and instant data about the coronavirus (2019-nCoV) on its online website during the pandemic. In addition to the daily case status of the virus, expert comments, general evaluations, and articles were provided under many categories, such as sectoral effects. Therefore, it is important to analyze the contents of the Nytimes.com coronavirus (2019-nCoV) -related articles and reveal the relationships between emerging concepts. By analyzing the Nytimes.com unstructured textual data with artificial intelligence (AI) methods, how mainstream media pandemic discourse has evolved can be explored.

The methods used in the analysis of article content are transformed by digitalization, particularly with the use of AI-based software. The findings can be quickly obtained, and the human factors can be minimized. AI software interfaces are used for these processes. With natural language processing (NLP), one of the sub-branches of AI, keywords related to the research topic are queried, and the content and metadata of the articles published with these queried keywords are converted into a dataset. These resulting articles are converted into datasets that are automatically classified by going through various text processes. As a result of this classification, various findings are retrieved. In addition, by creating analytical models suitable for the research problem, the data are instantly evaluated, and the results are visualized. These processes are short, unlike traditional methods. It provides important information regarding textual data [8]. The most important contribution of the results of the analysis performed with NLP is that it reveals the relationships, which can provide further understanding to researchers. It also helps to discover new insights and open the “black box”. We used NLP because supervised and unsupervised learning, and particularly deep learning, are now widely used for modeling human language, but there is still a need for syntactic and semantic understanding, as well as domain expertise. These aspects are not necessarily present in the machine learning approaches that are currently in use. NLP is significant because it helps resolve ambiguity in language and offers a helpful quantitative structure to the data for applications that come after it, such as voice recognition or text analytics. Both benefits are vital for downstream applications. In this study, a similar method was used to analyze the textual contents of the coronavirus (2019-nCoV)-related articles on Nytimes.com. First, the descriptive statistics of the textual dataset were defined and sentiment measurements of the published content were conducted. Then, using Leximancer concept maps [9], we explained concepts and their relationships by visualizing them to identify pandemic discourse.

In conclusion, we believe that our work makes four primary contributions to the field, as follows:

1. We gathered information about how the mainstream media (Nytimes.com) sets the agenda for the community.
2. To evaluate the debate around the pandemic, we used VADER’s sentiment analysis. The findings provided us with the distance and closeness positions of a few topics connected to the pandemic discourse on Nytimes.com, disclosed in accordance with the conceptual definitions of those themes.
3. Using our dataset, we performed text classification, for which we created four alternative machine learning models. In these analysis results, the algorithm that demonstrated the best performance has been determined.

- Concerning crisis communication monitoring, we have presented an overview of the emotions evoked and topics covered by Nytimes.com's pandemic-related (2019-nCOV) coverage, which should be useful.

The remainder of this article is organized as follows: The literature related to our research is examined in Section 2, the proposed research method and data are reviewed in Section 3, and the results and evaluations of our study are examined in Section 4. The discussion and implications are stated in Section 5, and finally, the conclusions are presented in Section 6.

2. Related Works

Coronavirus (2019-nCOV) has been the most important agenda item for society and individuals in recent years. Studies have examined news in the mainstream media to understand and analyze the impact of coronavirus (2019-nCOV). The influence of national governments on news sources during the pandemic and the impact of the coronavirus (2019-nCOV) news on the tourism industry were discussed [10]. Exposure to news of coronavirus (2019-nCOV) in the mainstream media [11], as well as the effects of its news on consumer behavior and consumption habits [12], were examined. Van et al. evaluated the results of the use of coronavirus (2019-nCOV) in the news media [13]. Studies have been conducted to investigate the sectoral and social effects of the news media. These studies have achieved important results over the pandemic period. Research that analyzes public discourse during the pandemic often focuses on social media discourse [14–19]. These studies were combined with behavioral models and analytical frameworks to analyze public priorities and concerns [20].

2.1. Theoretical-Related Works

As in the agenda-setting theory [21], the dominant group, which has the power to control the mass media, also determines what people will discuss. Thus, the community agenda is formed [22]. Media giants publish news that is in line with their own worldviews and do not consider other views as important. The news is presented to the audience in a certain order. People tend to consider the news presented first as important, so the information given first always sets the agenda of people and society [23]. Developments in the theory also reveal new directions in research [22]. During the pandemic period, studies were conducted based on this theory, especially analyzing the coronavirus (2019-nCOV) discourses [24–27] of social media platforms. The Nytimes.com pandemic discourse will contribute to these findings because Nytimes.com has a prominent place in setting the agenda of society. Popular daily publications such as The New York Times and The Washington Post are “agenda setters” within the United States Media. These broadcasts then have a direct impact on local newspapers and television networks [28].

2.2. Practical-Related Works

Significant developments in the field of artificial intelligence (AI) have enabled these analyses to be performed. The ability to easily obtain data with various AI software applications and purposefully analyze them with NLP has led to the automation of research methods and a significantly shorter timeline for processes. The development of these methods has supplied the opportunity to perform content analysis used in qualitative research in a reliable way. Texts converted into numerical data with natural language processing have created the opportunity to make predictions and inferences using machine learning (ML) algorithms. NLP, a subfield of artificial intelligence (AI), is used in the analysis of unstructured textual data [29]. Software interfaces, such as Leximancer [30] and Orange3 [31], are used to perform NLP operations. Using these applications, large amounts of qualitative and unstructured data, consisting of text, are automatically classified, analyzed, and significant insights are identified. In this study, a dataset was created by collecting coronavirus (2019-nCOV)-related articles published on the Nytimes.com website.

Leximancer, an AI-based NLP program, was used for content and sentiment analysis by passing the data through the steps of the working model.

Leximancer detects key concepts (or words) within blocks of text in the complete dataset based on their similarity and association with other words. The glossary generated and classified automatically for each set of data was created using machine learning. Classified texts are indexed as main themes, concepts, and compound concepts. Leximancer is also a computational content analysis tool for qualitative data analysis that calculates the frequency of concepts, their associations, and their proximity to each other [9,32]. The conceptual structure of articles in this study's dataset was synthesized, indexed, measured, and visualized as in earlier studies of a similar nature. It has been used in many academic studies [33–36]. In this study, Leximancer 5.0 was used to analyze “coronavirus” articles on Nytimes.com.

Sentiment analysis is a classification method used in the perceptual analysis of unstructured textual contents [37]. There are various analytical methods for NLP. The most successful among these methods is the valence-aware dictionary for sentiment reasoning (VADER), which is available in the Natural Language Toolkit (NLTK) library. VADER is sensitive to both the polarity (positive or negative) and the intensity (strength) of perception [38]. VADER is directly applied to unlabeled text data [39]. The sentiment classification (positive or negative) of the text was performed by directly applying the VADER that was coded in Python.

There have been numerous studies analyzing pandemic events on social media using machine learning techniques and natural language processing [40]. One study used five machine learning algorithms (decision tree, logistic regression, k-nearest neighbors, random forest, and support vector machine) based on a historically tagged dataset of coronavirus (2019-nCoV) tweets to build a sensitivity classifier [41]. Twitter offers COVIDSensing, a real-time tool where they use topic modeling and sentiment analysis to analyze socio-economic issues related to coronavirus (2019-nCoV) on Really Simple Syndication and Telegram [42]. One study detected “flu” on Twitter with a support vector machine to classify negative and positive flu tweets [43]. Andreadis et al. [44] analyzed tweets about coronavirus (2019-nCoV) in Italy. They used logistic regression and random forests to classify fake news or misinformation. Apart from tweets, the authors of [45] combined random forest, stochastic gradient descent, and a logistic regression decision and created a predictive model for the retweetability of tweets posted regarding coronavirus (2019-nCoV).

In one study, a novel mesoscopic text representation was proposed to understand what happens at the mid-sentence level [46]. Additionally, they defined and introduced a network-based representation of textual data that separates the synchronic properties of a text from its diachronic properties. Mansoor et al. examined the sentiment of tweets related to coronavirus (2019-nCoV) and how sentiment in different countries has changed over time. They proposed a framework to monitor and analyze sentiment towards the novel coronavirus (2019-nCoV) over time on Twitter and sentiment towards Work from Home and Online Learning [47]. Another study introduced a new method of training word embeddings with information about context to induce word senses in the word sense induction task. Results show that word embeddings can be used to address limitations in a manually annotated corpus [48].

The above analyses focused on Twitter data. There are no studies analyzing Nytimes.com data. In our study, a sentiment measurement was performed using the VADER method. The data were then evaluated with ML models (random forest, naïve Bayes, support vector machine, multilayer perceptron, Bert, and Electra).

3. Data and Methods

We registered for a New York Times Developer Account membership to collect research data. Using the Access Token key given by the Nytimes.com API, we searched for “coronavirus” as a keyword for three different periods (first, mid, and last) with the Orange3 Text Data Mining [31] feature, as seen in Figure 1. For each query, we obtained

1000 text records. A total of 3000 text items were generated for the three queries. We converted this text data into a dataset and saved it in .txt and .csv formats separately for each period.

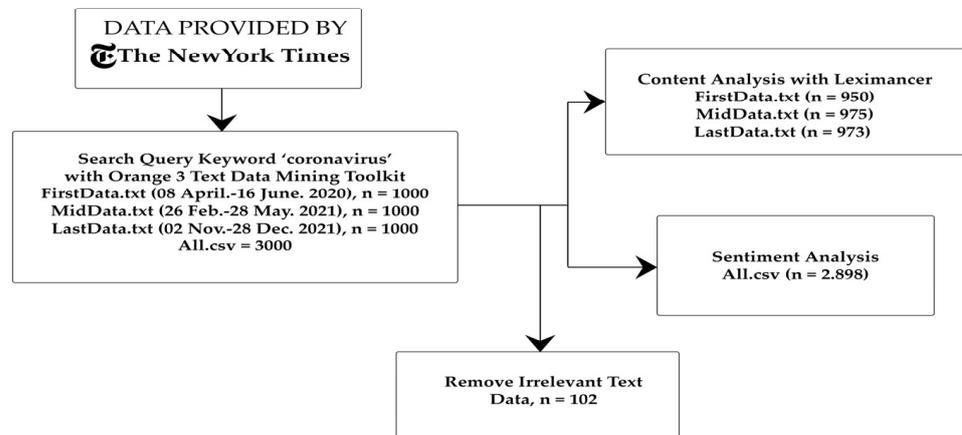


Figure 1. Structure and process of the Nytimes.com discourse analysis.

The contents of our dataset were as follows: We chose a time close to March 2020, the day the WHO declared the first case. First-term content ranged from 08 April to 16 June 2020, mid-term content from 26 February to 28 May 2021, and last-term content covered publications from 02 November to 28 December 2021. We chose the date ranges of the data by paying attention to the course of the pandemic, as shown in Figure 1. Each period holds approximately three months of text data. We removed duplicate data that could have influenced the analysis results. In this way, we have removed 102 rows of data from the analysis. In Figure 1, the datasets are sized as follows: first term ($n = 950$), mid-term ($n = 975$), and last term ($n = 973$).

For the descriptive analysis, we recorded the data separately for each period, merged the data into a single dataset (All.csv, $n = 2898$), and obtained the analysis results, as shown in Figure 2 for the Nytimes.com text dataset descriptive analysis. We then subjected this dataset to sentiment analysis. Finally, we created a project using the trial version of Leximancer software and performed a content analysis for our three-term data, separately.

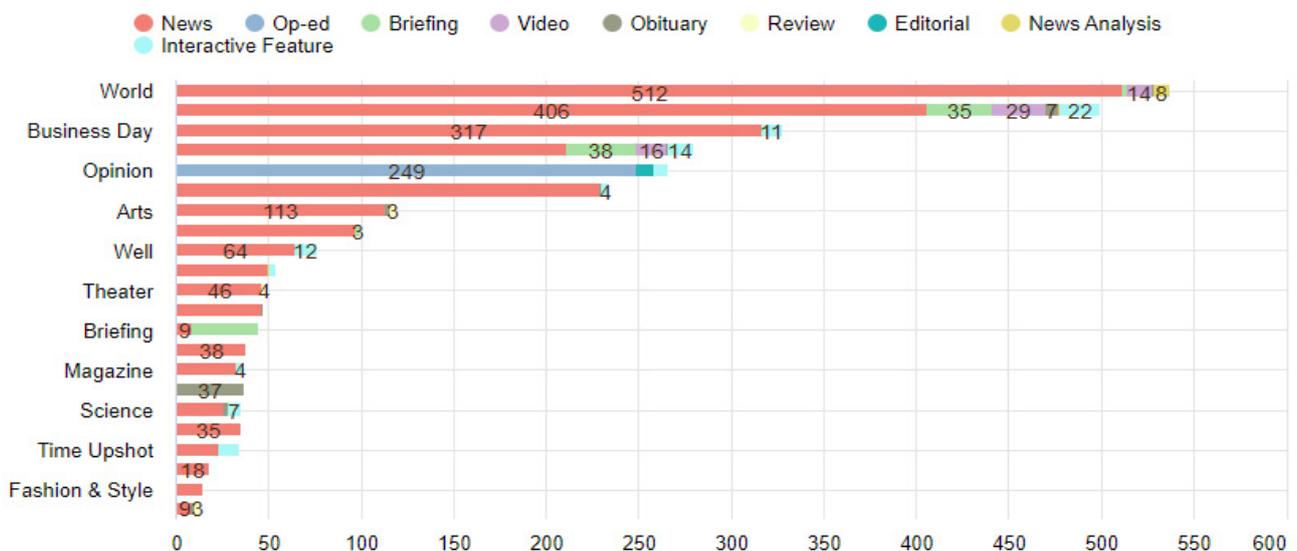


Figure 2. Nytimes.com article dataset box plot for descriptive analysis.

4. Results

For descriptive analysis, we combined the data in a single dataset (All.csv, $n = 2898$), which was recorded separately for each period. According to the textual dataset of the three periods, articles were mostly in the “World” section, and the least in the “Books” section. The textual dataset’s article types were as follows: news ($n = 2297$), op-ed ($n = 249$), briefing ($n = 114$), interactive feature ($n = 96$), video ($n = 59$), obituary ($n = 54$), review ($n = 10$), editorial ($n = 10$), and news analysis ($n = 9$), ($\chi^2: 24,000.00$, $p = 0.0001$, $dof = 64$).

The distribution of article types and section variables is shown in Figure 2 ($\chi^2: 7830.46$, $p = 0.0001$, $dof = 304$).

4.1. Sentiment Analysis and Results

We measured the sentiment scores using VADER, which is the most suitable for our dataset and was coded using Python. VADER is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. A sentiment lexicon is a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative. VADER has been found to be quite successful when dealing with social media texts, Nytimes editorials, movie reviews, and product reviews. This is because VADER not only obtains the positivity and negativity scores but also explains how positive or negative a sentiment is. Additionally, VADER sentiment analysis relies on a dictionary that maps lexical features to emotion intensities, known as sentiment scores. The sentiment score of a text can be obtained by summing up the intensity of each word in the text [38,49].

In comparison to more conventional approaches to sentiment analysis, VADER offers a number of benefits, including the following:

- It performs very well on text, similar to that seen on social networking platforms, while easily generalizing to a variety of other fields.
- It is developed using a generalizable, valence-based, human-curated gold standard sentiment lexicon, yet it does not need any training data.
- It has a speed that allows it to be utilized online with streaming data, and it does not suffer from a speed–performance tradeoff to a significant degree.

Besides these advantages, VADER has some disadvantages, such as:

- Analysis is language-specific.
- Discriminating jargon, nomenclature, memes, or turns of phrase may not be recognized.

Table 1 shows that 1157 of the 2898 textual contents related to the coronavirus (2019-nCoV) on Nytimes.com were negative sentiments (40%), and 1741 were positive sentiments (60%). These findings of negative content can be interpreted as being quite high compared to the normal agenda.

Table 1. Sentiment analysis results.

Row Labels	Count of Text	Percentage (%)
Positive	1741	60%
Negative	1157	40%
Total	2898	100%

4.2. Evaluation

In the supervised sentiment analysis, we used ML models and transformer-based approaches. The reason we chose these models was to compare the performance of traditional machine learning algorithms and transformer-based approaches.

We tested a variety of ML models and examined how well they performed relative to the performance of our trained model utilizing the text model as the baseline. The assessment was conducted in accordance with the usual classification metrics, which included precision, recall, F1-score, and accuracy. The evaluation results obtained from

naive Bayes, random forest, support vector machine, and multilayer perceptron, as well as transformer-based approaches Bert and Electra, classification models that were trained using our data and text embeddings, respectively, are shown in Tables 2 and 3, respectively.

Table 2. Calculations of accuracy, recall, and F1-score as macro-averages for evaluations using first data embeddings.

Method	Precision	Recall	F1-Score	Accuracy
Random Forest	0.64	0.72	0.64	0.78
Naïve Bayes	0.58	0.60	0.59	0.86
Support Vector Machine	0.59	0.61	0.60	0.85
Multilayer Perceptron	0.73	0.74	0.73	0.86
Bert	0.66	0.65	0.56	0.81
Electra	0.70	0.70	0.62	0.84

Table 3. Calculations of accuracy, recall, and F1-score as macro-averages for baseline evaluations using second data embeddings.

Method	Precision	Recall	F1-Score	Accuracy
Random Forest	0.62	0.70	0.62	0.76
Naïve Bayes	0.56	0.58	0.57	0.84
Support Vector Machine	0.57	0.59	0.58	0.83
Multilayer Perceptron	0.71	0.72	0.71	0.84
Bert	0.65	0.67	0.62	0.83
Electra	0.69	0.68	0.59	0.82

According to the findings of the evaluation, our model performed better than the second model for each ML model (the accuracy of both was the same for naive Bayes and multilayer perceptron, the only ML models for which this was not the case, and transformer-based approaches Bert and Electra). When comparing ML models that were employed in the supervised task of sentiment analysis, multilayer perceptron earned the greatest scores of all evaluation metrics in both representation models. This was the case regardless of which model was chosen. Specifically, the F1-score for the first model was 0.73, while the F1-score for the second model was 0.71. Despite having higher performance values than the general average, the transformer-based approaches Bert and Electra were unable to achieve the highest model scores. To determine which algorithm produces the best results, class-wise performance measurements were performed across positive and negative classes. Table 4 shows the class-wise performance of the ML models and transformer-based approaches. Although the number of favorable comments was much smaller than the number of comments in other classes, all the trained models have the potential to perform more effectively with a dataset that is more evenly distributed.

Table 4. Class-wise performance of the ML models and transformer-based approaches.

Class	RF	NB	SVM	MP	Bert	Electra
Positive	74.5	86.1	85.7	87.3	84.6	81.2
Negative	78.2	82.6	81.3	82.5	82.2	84.3

Electra, which is one of the transformer-based techniques, earned the maximum performance for the negative class, while multilayer perceptron achieved the best performance for the positive class (87.3%).

4.3. Leximancer Content Analysis and Results

When the Leximancer analysis process was mapped, concepts were clustered into higher-level “themes”. Concepts that often appear together in the same text strongly attract

4.3.4. Leximancer Content Analysis Results for Three Periods, Themes and Concepts

Three periods (first, mid, and last) are listed in Table 5, to summarize the Nytimes.com coronavirus (2019-nCOV) discourse between the analyzed periods of themes and concepts that appeared from the Leximancer analysis results. Thus, a comparison of the above visualized findings was provided to better understand the periodic change of discourse.

Table 5. Leximancer content analysis results for the three periods, themes and concepts.

First-Term Content (8 April–16 June 2020)		Mid-Term Content (26 February–28 May 2021)		Last-Term Content (2 November–28 December 2021)	
Themes	Concepts	Themes	Concepts	Themes	Concepts
health	country, officials, United States, states, public, spread, announced, outbreak, government, infections, million	Coronavirus	people, United States, country, months, India, world, Washington, millions, virus, administration announced, fully, Centers for Disease Control,	Coronavirus	vaccinated, announced, week, positive, testing, months, public
Coronavirus	pandemic, people, cases, Washington, world, virus	vaccinated	Americans, time, federal, public, masks, government, summer Covid, health, officials, week, vaccines, weeks Johnson, vaccination, doses, million, number, rare	variant	Omicron, cases, country, surge, restrictions, travel, spread, infections
months	month, March, city, lockdown	vaccine	restrictions, month, past, American, travel	pandemic	people, United States, world, countries
home	year, weeks, New York	cases	night, economic	covid	health, officials, government, virus
time	New York City, down, work	pandemic	others, series	vaccine	booster, shots, federal, administration, Biden mandate, vaccination, employees
day	week, care, days	summer	reopening, home, students	workers	Christmas
spread	outbreak, businesses	died	during, return	day	vaccines, Americans, rising, case
during	workers	school	reopening	down	children, city, company
died	series, others	year	news	children	policy
lockdown	lockdown	family	take	weeks	Prices, time
Covid	long	days	students	inflation	nearly
businesses	economic, restrictions, company	shot	hospital	down	
President Trump	Washington	students	life	long	Omicron
travel	travel	patients		unvaccinated	During, wave
latest	latest	office		school	Percent

5. Discussion and Implications

Leximancer analysis results are listed in Table 4 for the three periods. The concepts of coronavirus (2019-nCOV), restrictions, spread, people, children, school, and travel were included in the Nytimes.com discourse of coronavirus (2019-nCOV). In addition, the theme of coronavirus (2019-nCOV), which took place in all periods, took the second place after the first period. There was no discourse about vaccination in the first term, and it was revealed that vaccination occupied a large place in discourses in other periods. However, the last term showed content in the form of vaccinated and unvaccinated. Concepts related to time (day, day, week, time, latest, year) took place in all periods. Concepts related to official institutions were used in the contents of the articles, as an institution (government, federal, Centers for Disease Control, administration, authorities), as a title (President), and as a place (Washington, country, countries) in discourses in every period. This result shows that Nytimes.com has published the statements of these officials. Another finding is that the content related to death was included in the first two periods. Although there has been no

decrease in the number of deaths from the coronavirus (2019-nCoV), the concept of “death” has not taken place recently. From this, it can be concluded that virus-related deaths have become normalized. It can also be said that the feeling of death in the first two periods was the anchor of convincing people to get vaccinated.

From the findings related to the results in Table 4 and the studies in the literature on coronavirus (2019-nCoV), the established consensus is that the mainstream news media do not directly spread lies and publish official statements as endorsements [50]. Nytimes.com content analysis results confirmed that this journalism is close to official statements and confirms political decisions. Despite research and articles on the impact of restrictions on human behavior [51,52], the concept of “lockdown” only took place in the first-term discourse. Although the restrictions continued in other periods, the content did not take place in the mid-term and long-term periods. It could not pass the concepts of “vaccine and vaccinated” in terms of frequency of use. Jia and Lu claimed in some of their media columns that China’s handling of coronavirus (2019-nCoV) created negative public opinion against China [53]. However, in Nytimes.com articles, apart from the USA, only India was mentioned as a country name. China, which is thought to be the origin of the coronavirus (2019-nCoV) and has the densest population, was not mentioned on Nytimes.com.

In studies conducted during the initial period of the coronavirus (2019-nCoV) crisis, it has been concluded that the pandemic will have serious economic consequences worldwide and will affect every country [54], and that it has a noticeable effect on global economic growth [55]. As in the literature, the contents of Nytimes.com were treated as “economic” and “business” concepts in the first two periods. As in other recent studies [56,57], “inflation and price” emerged in this study. Even though the pandemic has negatively affected most areas, the concepts of “coronavirus (2019-nCoV)” and “travel” have come together in every period. We can say that the companies that provide services in this field are among the sectors that have suffered the most, suffering large financial losses due to travel restrictions. Contrary to this finding, there was no discourse related to the most profitable sectors during the pandemic. Furthermore, about 60 different variants have been detected and reported in the United Kingdom as of 13 December 2019 regarding the coronavirus (2019-nCoV) [58], but the concept of “variant” was only found on Nytimes.com in the last-term period, referred to as “Omicron”. During the H1N1 “Swine Flu” pandemic, a mandatory employee influenza vaccination policy was introduced for the 2009–2010 flu season. Mandatory influenza vaccination among healthcare workers (HCWs) is part of a mandatory seasonal flu vaccination program [59,60]. In particular, policies have been formed regarding the compulsory influenza vaccination of health workers at the Children’s Hospitals [61]. When we analyzed recent articles on Nytimes.com, articles related to the mandatory vaccination of duty employees were included in the coronavirus (2019-nCoV) crisis, as in past practices. Vaccine- and vaccination-related contents were not included in the first-term discourse but became a full agenda in the other two periods. In addition, although the World Health Organization approved the Pfizer-BioNTech, Oxford-AstraZeneca, Moderna, and Janssen vaccines for emergency use, only the Johnson vaccine came to the fore in the article content.

6. Conclusions

According to the findings of the Nytimes.com content analysis regarding coronavirus (2019-nCoV), it has had many behavioral effects, such as inducing fear and anxiety in people, and it can be concluded that the discourse around coronavirus (2019-nCoV), which has caused great changes and results such as death, restrictions, and mask use, did not go beyond a total of 15 main themes and about 100 concepts. The results of content and sentiment analysis supported this inference. Considering the media dependency of society, this result is important. It can be said that the agenda-setting of society does not go beyond the traditional discourse due to the tendency of individuals to use newspapers and news websites to obtain information. In Leximancer analysis results, it was seen that the news contents were conceptualized as “virus”, “spread”, “millions”, “people”, and “died”,

and this is supported by official discourse. It can be concluded that related concepts are agenda-setting.

There are periodic differences in the coronavirus (2019-nCoV) pandemic discourse on Nytimes.com. As a result of the Leximancer analysis, themes and concepts were found to run periodically parallel to the course of the pandemic. It can be concluded that the influence on the economy and restrictions on businesses, which took place in the first period, resulted in inflation and price increases in the last period. While there was no discourse about vaccination in the first period, vaccination discourse took place extensively in the last two. While the discourse in the contents of the first period was more formal and restrictive, it was seen that the restrictions decreased with the reopening of schools in the last two periods, and the discourse on using masks and getting vaccinated was included in the mid-term period. In Nytimes.com coronavirus (2019-nCoV) discourses, partial closeness to the literature and the pandemic process showed that the discourse was in parallel with official policies. Nytimes.com public discourse can support the horse-racing journalism view, with journalism being too close to official statements and approving political decisions.

The theoretical contribution of this research was to show how society's agenda was created during the pandemic by analyzing Nytimes.com text contents using NLP methods. It can help break the "black box" of mainstream media. This contribution is also important considering the ability of the media to manipulate individuals through behavioral patterns such as fear and anxiety. It can reveal the importance of discourse in restraining society and persuading various practices. On the other hand, as a practical contribution, in this study, four distinct and widely used ML models and transformer-based approaches were used. The reason we chose these models was to compare the performance of traditional machine learning algorithms and transformer-based approaches. At this point, the text data that have been gathered were separated into two distinct datasets. To determine which algorithm produced the best results, class-wise performance measurements were performed across positive and negative classes. Electra, which is one of the transformer-based techniques, showed the best performance for the negative class, while multilayer perceptron achieved the best performance for the positive class (87.3%).

Future studies, not limited to Nytimes.com, may include discourses from other mainstream media outlets. In addition, the conceptual relationships that emerge in the findings and concern other disciplines may be the subject of future research.

The limitations of the study are that the Nytimes.com API only supplies access to 1000 articles at a time, so the full content of the pandemic discourse could not be accessed. It is also limited in that we analyzed only one mainstream media outlet.

This research has proven the possibilities afforded by the convenience and usefulness of natural language processing methods, which can process a large amount of textual data and provide insights into the sentiment and topics of the observed texts. When it comes to the task of analyzing public opinion and attitudes toward a variety of coronavirus (2019-nCoV)-related themes, natural language processing can complement the achievements of traditional approaches that are used in research in the fields of the humanities and social sciences. In this way, natural language processing can complement the achievements of traditional approaches used in research. When it comes to the duty of monitoring online discussion surrounding the coronavirus (2019-nCoV) pandemic, we feel that our work contributes to the pursuit of the increasing social media research that is now being conducted.

Author Contributions: Conceptualization, S.T.; methodology, S.T.; software, S.T.; validation, S.T. and B.S.; formal analysis, B.S.; investigation, B.S. and Y.S.B.; resources, S.T. and Y.S.B.; data curation, S.T.; writing—original draft preparation, S.T. and Y.S.B.; writing—review and editing, S.T. and B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used in this study can be obtained by contacting the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Caduff, C. What Went Wrong: Corona and the World after the Full Stop. *Med. Anthropol. Q.* **2020**, *34*, 467–487. [CrossRef]
2. Cheval, S.; Mihai Adamescu, C.; Georgiadis, T.; Herrnegger, M.; Pitaric, A.; Legates, D.R. Observed and Potential Impacts of the COVID-19 Pandemic on the Environment. *IJERPH* **2020**, *17*, 4140. [CrossRef]
3. Wiederhold, B.K. Using Social Media to Our Advantage: Alleviating Anxiety During a Pandemic. *Cyberpsychol. Behav. Soc. Netw.* **2020**, *23*, 197–198. [CrossRef]
4. Viehmann, C.; Ziegele, M.; Quiring, O. Communication, Cohesion, and Corona: The Impact of People's Use of Different Information Sources on Their Sense of Societal Cohesion in Times of Crises. *J. Stud.* **2022**, *23*, 629–649. [CrossRef]
5. Newman, N. *Reuters Institute Digital News Report 2020*; Reuters Institute for the study of Journalism: Oxford, England, 2020.
6. Kim, S.J.; Wang, X.; Malthouse, E.C. Digital News Readership and Subscription in the United States during COVID-19: A Longitudinal Analysis of Clickstream and Subscription Data from a Local News Site. *Digit. J.* **2022**, *10*, 1015–1036. [CrossRef]
7. Andersen, K.; Shehata, A.; Andersson, D. Alternative News Orientation and Trust in Mainstream Media: A Longitudinal Audience Perspective. *Digit. J.* **2021**, 1–20. [CrossRef]
8. Lucy, L.; Demszky, D.; Bromley, P.; Jurafsky, D. Content Analysis of Textbooks via Natural Language Processing: Findings on Gender, Race, and Ethnicity in Texas U.S. History Textbooks. *AERA Open* **2020**, *6*, 233285842094031. [CrossRef]
9. Leximancer User Guide. 2022, p. 149. Available online: <https://www.doc.leximancer.com/> (accessed on 20 March 2022).
10. Chen, H.; Huang, X.; Li, Z. A Content Analysis of Chinese News Coverage on COVID-19 and Tourism. *Curr. Issues Tour.* **2022**, *25*, 198–205. [CrossRef]
11. Olagoke, A.A.; Olagoke, O.O.; Hughes, A.M. Exposure to Coronavirus News on Mainstream Media: The Role of Risk Perceptions and Depression. *Br. J. Health Psychol.* **2020**, *25*, 865–874. [CrossRef]
12. Cruz-Cárdenas, J.; Zabelina, E.; Guadalupe-Lanas, J.; Palacio-Fierro, A.; Ramos-Galarza, C. COVID-19, Consumer Behavior, Technology, and Society: A Literature Review and Bibliometric Analysis. *Technol. Forecast. Soc. Change* **2021**, *173*, 121179. [CrossRef]
13. Van Aelst, P.; Toth, F.; Castro, L.; Štětka, V.; Vreese, C.d.; Aalberg, T.; Cardenal, A.S.; Corbu, N.; Esser, F.; Hopmann, D.N.; et al. Does a Crisis Change News Habits? A Comparative Study of the Effects of COVID-19 on News Media Use in 17 European Countries. *Digit. J.* **2021**, *9*, 1208–1238. [CrossRef]
14. Kurten, S.; Beullens, K. #Coronavirus: Monitoring the Belgian Twitter Discourse on the Severe Acute Respiratory Syndrome Coronavirus 2 Pandemic. *Cyberpsychol. Behav. Soc. Netw.* **2021**, *24*, 117–122. [CrossRef]
15. Ellerich-Groppe, N.; Pfaller, L.; Schweda, M. Young for Old—Old for Young? Ethical Perspectives on Intergenerational Solidarity and Responsibility in Public Discourses on COVID-19. *Eur. J. Ageing* **2021**, *18*, 159–171. [CrossRef]
16. Ayalon, L.; Chasteen, A.; Diehl, M.; Levy, B.R.; Neupert, S.D.; Rothermund, K.; Tesch-Römer, C.; Wahl, H.-W. Aging in Times of the COVID-19 Pandemic: Avoiding Ageism and Fostering Intergenerational Solidarity. *J. Gerontol. Ser. B* **2021**, *76*, e49–e52. [CrossRef]
17. Xiang, X.; Lu, X.; Halavanau, A.; Xue, J.; Sun, Y.; Lai, P.H.L.; Wu, Z. Modern Senicide in the Face of a Pandemic: An Examination of Public Discourse and Sentiment About Older Adults and COVID-19 Using Machine Learning. *J. Gerontol. Ser. B* **2021**, *76*, e190–e200. [CrossRef]
18. Pascual-Ferrá, P.; Alperstein, N.; Barnett, D.J. Social Network Analysis of COVID-19 Public Discourse on Twitter: Implications for Risk Communication. *Disaster Med. Public Health Prep.* **2022**, *16*, 561–569. [CrossRef]
19. Xue, J.; Chen, J.; Chen, C.; Zheng, C.; Li, S.; Zhu, T. Public Discourse and Sentiment during the COVID 19 Pandemic: Using Latent Dirichlet Allocation for Topic Modeling on Twitter. *PLoS ONE* **2020**, *15*, e0239441. [CrossRef]
20. Habib, M.A.; Anik, M.A.H. Impacts of COVID-19 on Transport Modes and Mobility Behavior: Analysis of Public Discourse in Twitter. *Transp. Res. Rec.* **2021**, *2*, 036119812110299. [CrossRef]
21. Agenda-Setting Theory. 2022. Available online: <https://wikipedia> (accessed on 20 March 2022).
22. McCombs, M.E.; Shaw, D.L.; Weaver, D.H. New Directions in Agenda-Setting Theory and Research. *Mass Commun. Soc.* **2014**, *17*, 781–802. [CrossRef]
23. Littlejohn, S.W.; Foss, K.A. *Theories of Human Communication*, 10th ed.; Waveland Press: Long Grove, IL, USA, 2010; ISBN 978-1-4786-0939-1.
24. Dai, Y.; Li, Y.; Cheng, C.-Y.; Zhao, H.; Meng, T. Government-Led or Public-Led? Chinese Policy Agenda Setting during the COVID-19 Pandemic. *J. Comp. Policy Anal. Res. Pract.* **2021**, *23*, 157–175. [CrossRef]
25. Meutia, I.F.; Sujadmiko, B.; Yulianti, D.; Putra, K.A.; Aini, S.N. The Agenda Setting Policy for Hajj and Umrah in Post Pandemic. In Proceedings of the 2nd International Indonesia Conference on Interdisciplinary Studies (IICIS 2021), Amsterdam, The Netherlands, 28 October 2021; pp. 32–37.

26. Liu, K.; Geng, X.; Liu, X. The Application of Network Agenda Setting Model during the COVID-19 Pandemic Based on Latent Dirichlet Allocation Topic Modeling. *Front. Psychol.* **2022**, *13*, 954576. [[CrossRef](#)] [[PubMed](#)]
27. Wang, Q. Using Social Media for Agenda Setting in Chinese Government's Communications during the 2020 COVID-19 Pandemic. *J. Commun. Inq.* **2022**, *46*, 01968599221105099. [[CrossRef](#)]
28. Wikipedia Agenda-Setting Theory-Wikipedia. Available online: https://en.wikipedia.org/wiki/Agenda-setting_theory (accessed on 23 November 2022).
29. Manning, C.; Surdeanu, M.; Bauer, J.; Finkel, J.; Bethard, S.; McClosky, D. The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, MD, USA, 23–24 June 2014; pp. 55–60.
30. Leximancer. Available online: <https://www.leximancer.com/> (accessed on 20 March 2022).
31. Orange Data Mining-Text Mining. Available online: <https://orangedatamining.com/workflows/Text-Mining/> (accessed on 20 March 2022).
32. Kanasa, D.H. An Introduction to Leximancer. Available online: <https://www.leximancer.com/> (accessed on 20 March 2022).
33. Angus, D.; Rintel, S.; Wiles, J. Making Sense of Big Text: A Visual-First Approach for Analysing Text Data Using Leximancer and Discursis. *Int. J. Soc. Res. Methodol.* **2013**, *16*, 261–267. [[CrossRef](#)]
34. Wilk, V.; Cripps, H.; Capatina, A.; Micu, A.; Micu, A.-E. The State of #digitalentrepreneurship: A Big Data Leximancer Analysis of Social Media Activity. *Int. Entrep. Manag. J.* **2021**, *17*, 1899–1916. [[CrossRef](#)]
35. Wilk, V.; Soutar, G.N.; Harrigan, P. Tackling Social Media Data Analysis: Comparing and Contrasting QSR NVivo and Leximancer. *QMR* **2019**, *22*, 94–113. [[CrossRef](#)]
36. Tunca, S.; Sezen, B.; Wilk, V. An Exploratory Content and Sentiment Analysis of The Guardian Metaverse Articles Using Leximancer and Natural Language Processing. *Cyberpsychol. Behav. Soc. Netw.* **2022**, *26*, 56–78.
37. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment Classification Using Machine Learning Techniques. *arXiv* **2002**, arXiv:cs/0205070.
38. Hutto, C.J.; Gilbert, E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In Proceedings of the International AAAI Conference on Web and Social Media, Ann Arbor, MI, USA, 1–4 June 2014.
39. Beri, A. Sentimental Analysis Using Vader. Interpretation and Classification of ... | by Aditya Beri | Towards Data Science. Available online: <https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664?gi=ee44b81a54cb> (accessed on 3 February 2022).
40. Qin, Z.; Ronchieri, E. Exploring Pandemics Events on Twitter by Using Sentiment Analysis and Topic Modelling. *Appl. Sci.* **2022**, *12*, 11924. [[CrossRef](#)]
41. Zhang, X.; Saleh, H.; Younis, E.M.G.; Sahal, R.; Ali, A.A. Predicting Coronavirus Pandemic in Real-Time Using Machine Learning and Big Data Streaming System. *Complexity* **2020**, *2020*, 6688912. [[CrossRef](#)]
42. Sepúlveda, A.; Periñán-Pascual, C.; Muñoz, A.; Martínez-España, R.; Hernández-Orallo, E.; Cecilia, J.M. COVIDSensing: Social Sensing Strategy for the Management of the COVID-19 Crisis. *Electronics* **2021**, *10*, 3157. [[CrossRef](#)]
43. Aramaki, E.; Maskawa, S.; Morita, M. Twitter Catches The Flu: Detecting Influenza Pandemics Using Twitter. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland, UK, 27–31 July 2011.
44. Andreadis, S.; Antzoulatos, G.; Mavropoulos, T.; Giannakeris, P.; Tzionis, G.; Pantelidis, N.; Ioannidis, K.; Karakostas, A.; Gialampoukidis, I.; Vrochidis, S.; et al. A Social Media Analytics Platform Visualising the Spread of COVID-19 in Italy via Exploitation of Automatically Geotagged Tweets. *Online Soc. Netw. Media* **2021**, *23*, 100134. [[CrossRef](#)] [[PubMed](#)]
45. Mahdikhani, M. Predicting the Popularity of Tweets by Analyzing Public Opinion and Emotions in Different Stages of COVID-19 Pandemic. *Int. J. Inf. Manag. Data Insights* **2022**, *2*, 100053. [[CrossRef](#)]
46. de Arruda, H.F.; Silva, F.N.; Marinho, V.Q.; Amancio, D.R.; Costa, L. da F. Representation of Texts as Complex Networks: A Mesoscopic Approach. *J. Complex Netw.* **2018**, *6*, 125–144. [[CrossRef](#)]
47. Mansoor, M.; Gurumurthy, K.; Prasad, V.R. Global Sentiment Analysis Of COVID-19 Tweets Over Time. *arXiv* **2020**, arXiv:2010.14234.
48. Corrêa, E.A.; Amancio, D.R. Word Sense Induction Using Word Embeddings and Community Detection in Complex Networks. *Phys. A Stat. Mech. Its Appl.* **2019**, *523*, 180–190. [[CrossRef](#)]
49. Pandey, S. Simplifying Sentiment Analysis Using VADER in Python (on Social Media Text) | by Parul Pandey | Analytics Vidhya | Medium. Available online: <https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f> (accessed on 25 December 2022).
50. Quandt, T.; Boberg, S.; Schatto-Eckrodt, T.; Frischlich, L. Pandemic News: Facebook Pages of Mainstream News Media and the Coronavirus Crisis—A Computational Content Analysis. *arXiv* **2020**, arXiv:2005.13290.
51. Ahmed, S.; Khaium, M.O.; Tazmeem, F. COVID-19 Lockdown in India Triggers a Rapid Rise in Suicides Due to the Alcohol Withdrawal Symptoms: Evidence from Media Reports. *Int. J. Soc. Psychiatry* **2020**, *66*, 827–829. [[CrossRef](#)]
52. Lemenager, T.; Neissner, M.; Koopmann, A.; Reinhard, I.; Georgiadou, E.; Müller, A.; Kiefer, F.; Hillemacher, T. COVID-19 Lockdown Restrictions and Online Media Consumption in Germany. *IJERPH* **2020**, *18*, 14. [[CrossRef](#)]
53. Jia, W.; Lu, F. US Media's Coverage of China's Handling of COVID-19: Playing the Role of the Fourth Branch of Government or the Fourth Estate? *Glob. Media China* **2021**, *6*, 8–23. [[CrossRef](#)]
54. Donthu, N.; Gustafsson, A. Effects of COVID-19 on Business and Research. *J. Bus. Res.* **2020**, *117*, 284–289. [[CrossRef](#)]

55. Crs, R. *Global Economic Effects of COVID-19*; Congressional Research Service: Washington, DC, USA, 2020; Volume 84, pp. 20–115.
56. Apergis, E.; Apergis, N. Inflation Expectations, Volatility and Covid-19: Evidence from the US Inflation Swap Rates. *Appl. Econ. Lett.* **2021**, *28*, 1327–1331. [[CrossRef](#)]
57. Coluccia, B.; Agnusdei, G.P.; Miglietta, P.P.; De Leo, F. Effects of COVID-19 on the Italian Agri-Food Supply and Value Chains. *Food Control* **2021**, *123*, 107839. [[CrossRef](#)] [[PubMed](#)]
58. Wise, J. COVID-19: New Coronavirus Variant Is Identified in UK. *BMJ* **2020**, *371*, m4857. [[CrossRef](#)] [[PubMed](#)]
59. Babcock, H.M.; Gemeinhart, N.; Jones, M.; Dunagan, W.C.; Woeltje, K.F. Mandatory Influenza Vaccination of Health Care Workers: Translating Policy to Practice. *Clin. Infect. Dis.* **2010**, *50*, 459–464. [[CrossRef](#)]
60. Hakim, H.; Gaur, A.H.; McCullers, J.A. Motivating Factors for High Rates of Influenza Vaccination among Healthcare Workers. *Vaccine* **2011**, *29*, 5963–5969. [[CrossRef](#)]
61. Douville, L.E.; Myers, A.; Jackson, M.A.; Lantos, J.D. Health Care Worker Knowledge, Attitudes, and Beliefs Regarding Mandatory Influenza Vaccination. *Arch. Pediatr. Adolesc. Med.* **2010**, *164*, 33–37. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.