PUBLIC PERCEPTION TOWARDS CHILDREN'S COVID-19 VACCINATION WITH NATURAL LANGUAGE PROCESSING

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF

ENGINEERING AND NATURAL SCIENCES

OF ISTANBUL MEDIPOL UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

HEALTHCARE SYSTEMS ENGINEERING

By

Elif Delice

June,2022

PUBLIC PERCEPTION TOWARDS CHILDREN’S COVID-19 VACCINATION WITH NATURAL LANGUAGE PROCESSING

By Elif Delice

3 June 2022

We certify that we have read this dissertation and that in our opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Master of Science.

Prof. Dr. Hakan Tozan (Advisor)

Dr. Öğr. Üyesi Mustafa Yağımlı

Dr. Öğr. Üyesi Cihan Bilge Kayasandık

Approved by the Graduate School of Engineering and Natural Sciences:

Prof. Dr. Yasemin Yüksel Durmaz

Director of the Graduate School of Engineering and Natural Sciences

I hereby declare that all information in this document has been obtained and presented in accordance with the academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Signature :

Name, Surname: ELİF DELİCE

ACKNOWLEDGEMENT

I would like to start by expressing my sincere thanks to my esteemed advisors, Prof. Dr. Hakan TOZAN and Assist. Prof. Dr. Melis Almula Karadayı, who contributed greatly to the emergence of this study, who always supported me at all times under all circumstances, also, who always roll a log for me throughout my academic progress and enlightened my path with their experiences. I am grateful to them for all my achievements so far. It is an honor for me to work with them.

I would also like to thank my dear mother Şehnaz TEKİN and my dear grandmother Yüksel TEKİN, who helped me get to where I am today, supported me with their endless love, stood by me in my hardest times, helped me become such a strong woman, and applauded my achievements. I am very happy to have grown up in such a family and I am proud to be their daughter. Without their patience and belief in me, it would not have been possible for me to achieve so much success today.

I would like to thank my dear friend Harun TEKİN, who always makes me smile, raises my motivation to the highest level, embraces every problem, sadness, and joy together, also, most importantly, I grew up together and reached today. I hope we will achieve many more successes together and applaud each other's happiness. Finally, I would like to express my special thanks to my dear friend Lütviye Özge POLATLI, who we found each other late even though we were in the same environment and shared all our troubles, who is no longer a friend but a sister to me.

Elif Delice

June, 2022

CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$\boldsymbol{k}$: Number of topics
$\boldsymbol{M}$: Number of texts
$\boldsymbol{z}$: Subject distribution of words
ω : Term
L : Text length
φ : Subject-word probability distribution
Θ : Text-topic probability distribution
α : Hyperparameter of Θ
β : Hyperparameter of φ
β_{ij} : The j -th word probability under the i

ABBREVIATIONS

BERT	**:** Bidirectional Encoder Statement from Transformers
CDC	**:** Centers for Disease Control and Prevention
COVID-19	**:** Novel Corona Virus Disease
IT	**:** Information Technologies
LDA	**:** Latent Dirichlet Allocation
LSA	**:** Latent Semantic Analysis
NLP	**:** Natural Language Processing
NMF	**:** Non-negative Matrix Factorization
PLSA	**:** Probabilistic Latent Semantic Analysis
SHAP	**:** Shapley Contribution Annotations
SVD	**:** Singular Value Decomposition
UK	**:** United Kingdom
USA	**:** United States
WHO	**:** World Health Organization

DOĞAL DİL İŞLEMEYLE ÇOCUKLARIN COVID-19 AŞISINA İLİŞKİN KAMUOYU ALGISI

ÖZET

Elif Delice

Sağlık Sistemleri Mühendisliği, Yüksek Lisans

Tez Danışmanı: Prof. Dr. Hakan Tozan

Eş Danışman: Dr. Öğr. Üye. Melis Almula Karadayı

Haziran, 2022

Korona virüs hastalığı 2019'un sonlarında Çin'in Wuhan kentinde kendini göstermiş, salgına dönüşerek küresel bir sorun haline gelmiş ve birçok ölüme sebebiyet vermiştir. Salgının yayılmasıyla geliştirilen stratejiler sonucunda, ebeveyn ve çocuklar hayatlarını derinden etkileyen kısıtlamalarla karşılaşmışlardır. Özellikle çocuklarda, sosyal hayatlarının kısıtlanması nedeniyle, depresyon, kaygının tetiklediği gerginlik gibi olumsuz psikolojik etkilere rastlanmıştır. COVID-19'a karşı geliştirilen aşılar pandemiyi sonlandırmanın veya hafifletmenin bir yolu olarak görülmektedir. Fakat, sürü bağışıklığının kazanılması adına Dünya nüfusunun yüksek çoğunluğunun aşıyı kabul etmesinin ve yaptırmasının gerekliliği vurgulanmaktadır. Bu noktada, hassas bir bünyeye sahip olan çocukların da normal hayatlarına geri dönebilmeleri adına aşılanmaları hakkında birçok tartışma ortaya çıkmış, kamuoyunun tepkisi sosyal medya aracılığıyla yankı bulmuştur. Yapılan detaylı literatür araması sonucunda, bilgimiz dahilinde, özellikle, ailelerin çocuklarına aşı yaptırma konusundaki yargılarını saptayan, metinlerin içindeki saklı anlamları ortaya çıkaran konu modelleme ve bireylerin duygusal durumlarını belirleyen duygu analizi tekniklerini birleştirip Twitter verilerinden destek alarak bir model ortaya koyan hiçbir çalışmaya rastlanmamış, genel olarak çalışmaların konu modelleme ve duygu analizi yöntemlerini ayrı ayrı kullanarak analizleri gerçekleştirdikleri ve bu yolla çıkarımlar yaptıkları gözlemlenmiştir.

Bu doğrultuda, literatürdeki eksikliğin gözlemlenmesiyle çalışma kapsamında, ebeveynlerin çocuklarına aşı yaptırma konusundaki algısını ortaya koyan, ana temaları çıkaran ve bu konular hakkında duygu değişimlerini saptayan bir model sunulması amaçlanmış, böylelikle, literatüre bu bağlamda katkı yapılması hedeflenmiştir. Ayrıca, uygulama sonucunda elde edilen çıktılarla, stratejilerini geliştirdikleri süreçte politika yapıcılara konu hakkında model sunarak ışık tutmak ve gelecek çalışmalara rehberlik yapmak hedeflenmiştir.

Çalışma kapsamında, Octoparse web kazıma aracının desteğiyle Twitter'dan salgının küresel sorun haline dönüştüğü ve aşılar hakkındaki tartışmaların yoğunlaştığı 1 Ağustos 2020-1 Ekim 2021 tarih çerçevesinde, tüm Dünya'yı ele alabilmek adına İngilizce olarak belirlenen anahtar kelimelerinin yardımıyla veriler çekilmiştir. Daha sonra, Doğal Dil İşleme'nin (NLP'nin) çatısı altında olan konu modelleme ve duygu analizi tekniklerini kullanarak, ebeveynlerin tutumları hakkında ana konular ile alt konular ortaya çıkarılmış, saptanan konular çerçevesinde duygu analizi gerçekleştirilerek, aşı tutumları saptanmıştır.

Yapılan çalışma sonucunda, “yaşa göre ilk doz aşılama ihtiyacı görüşü”, “ilk doz aşı etkinliği”, “okul çağındaki çocukların aşılanma ihtiyacı” ve “aşısız çocukların sadece maske koruması ile korunmasından kaynaklanan aşı ihtiyacı düşüncesi” olmak üzere dört konu kümesi belirlenmiştir. Belirlenen konu kümelerinin duygu analiziyle işlenmesiyle olumlu duyguların ağırlıkta olduğu, güven, beklenti ve korku olmak üzere üç duygunun öne çıktığı saptanmıştır.

Anahtar sözcükler: COVID-19, aşı, doğal dil işleme (NLP), konu modelleme, duygu analizi, web kazıma.

PUBLIC PERCEPTION TOWARDS CHILDREN'S COVID-19 VACCINATION WITH NATURAL LANGUAGE PROCESSING

ABSTRACT

Elif Delice

MSc in Healthcare Systems Engineering

Advisor: Prof. Dr. Hakan Tozan

Co-Advisor: Assoc. Prof. Melis Almula Karadayı

June, 2022

At the end of 2019, Coronavirus disease manifested itself in Wuhan, China, turned into an epidemic, became a global problem, and caused many deaths. As a result of the strategies developed with the spread of the epidemic, parents and children have faced restrictions that deeply affect their lives. Especially in children, negative psychological effects such as depression and tension triggered by anxiety have been encountered due to the restriction of their social lives. Vaccines developed against COVID-19 are seen as a way to end or mitigate the pandemic. However, it is emphasized that the majority of the world's population should accept and have the vaccine in order to gain herd immunity. At this point, many debates arose about the vaccination of children with a sensitive bodies so that they can return to their normal lives, and the public's reaction was echoed through social media. As a consequence of the detailed literature search, to the best of our knowledge, no study specifically determined the judgments of families about getting their children vaccinated, that combined topic modeling which reveals the hidden meanings in the texts, and sentiment analysis techniques, that determine the emotional states of individuals. supported by Twitter data, and presented a model, generally, it was seen that the studies carried out analyses using topic modeling, as well as, sentiment analysis methods separately and they made inferences in this way.

In this direction, by observing the deficiency in the literature, it is targeted to present a model that reveals the perception of parents about getting their children vaccinated, extracts the main themes, and determines the emotional changes about these topics, thus, it is intended to contribute to the literature in this area. Furthermore, it is hoped that the outputs acquired as a result of the implementation would shed light on the subject and guide future studies by presenting a model to the policymakers in the process of developing their strategies.

Scope of work, with the support of the Octoparse web scraping tool, data was extracted from Twitter with the help of keywords determined in English in order to address the whole world, within the framework of the date of August 1, 2020-October 1, 2021, when the epidemic turned into a global problem and the discussions about vaccines intensified. Then, using the topic modeling and sentiment analysis techniques under the umbrella of NLP, main, sub-topics about parents' attitudes were revealed, also vaccine perceptions and attitudes were detected by performing sentiment analysis within the framework of the identified subjects.

As a result, four topic clusters were determined: “the opinion of the need for the first dose of vaccination according to age”, “the effectiveness of the first dose of vaccine”, “the opinion of the need for vaccination of school-age children”, and “the need for vaccination arising from the protection of unvaccinated children with only mask protection”. With the processing of the determined topic clusters with sentiment analysis, it was determined that positive emotions were dominant, and three emotions, namely trust, expectation, and fear, came to the fore.

Keywords: COVID-19, vaccine, natural language processing (NLP), topic modeling, sentiment analysis, web scrapping.

CHAPTER 1

1. INTRODUCTION

At the end of 2019, the coronavirus, which originates from the severe acute respiratory syndrome, first emerged in Wuhan, China, spread rapidly to the world, and affected more than 220 countries, upending lives, thus reaching the epidemic status in 2020 and becoming a global threat [1]–[3]. There are currently more than 500 million cases of COVID-19 and more than 6 million deaths worldwide confirmed by WHO [4]. In order to alleviate the epidemic, basic measures were taken by governments, in this way, the public was tried to be protected in the development process of an effective vaccine. As a result, problems have emerged that deeply affect families with children. For children who cannot even perform their daily activities, the anxiety experienced by their parents created great pressure and negatively affected their psychology. It has been noted in this context that many studies in the literature try to determine the perception of the public about the developed coronavirus vaccines. The fact that the studies carried out, with a high majority, dealt with the adult population and preferred the survey method, which is a traditional way, brought the subject to the fore. The main motivation of the research is that this study can guide research to be carried out based on vaccination in children and policymakers in this process, by using the newest machine learning methods, with a comprehensive model presentation.

In many types of research, it was emphasized that the way to protect from the virus and minimize its negative effects is the rapid development of the vaccine. Therefore, great efforts are being made by governments to develop successful vaccines for the coronavirus and to promote vaccination. As a result of these efforts, more than 11 million vaccines have been administered worldwide [4]. However, although the widespread application of the vaccine is required in order to gain herd immunity, the prejudices and uneasiness of the public create obstacles in front of vaccination. While adults were still uneasy about getting themselves vaccinated, with issues such as reopening schools and stretching the measures, the vaccines to be administered to children became a problem, leading to heated debates in the public. It is vital for all of these reasons that policymakers understand public sentiment towards vaccines that will reduce the negative aspects of the epidemic [5].

Recently, social media platforms have become a place that provides information to many users on current issues and allows them to exchange ideas. In this context, Twitter has come to the fore, also, has become a tool for people around the world to share their thoughts and attitudes. During the epidemic, the public shared their reactions to both measures taken and the vaccines through these platforms. It has been suggested that observing the dissemination of such information as well as identifying the topics discussed will be beneficial for strategy developers in their future strategies and will enable them to determine the expectations of the public [6].

This study aims to identify the attitudes of parents about vaccinating their children in the time frame of 1 August 2020-1 November 2021, when the subject is most discussed, by addressing tweets published around the world containing certain keywords in the English language, and to analyze the characteristics of these attitudes. In the tweets, topic modeling and sentiment analysis framework will be put forward in order to identify the main topics, as well as to raise public awareness about vaccinating children. The first chapter provides a general introduction to the subject, while the second part conducts a complete evaluation of the literature. In the third chapter, the methodology was explained in detail, and application steps were shown. In the fourth chapter, the outputs of the model were discussed. And lastly, in the fifth chapter, the study was concluded and suggestions about future works were given.

1.1. COVID-19 in the World

The coronavirus disease has spread rapidly since December 2019, when it first appeared in Wuhan, China, and has turned into an epidemic, leaving many negative effects on the whole world. Starting in China, it infected more than 500 million people in more than 220 countries, thus causing more than 6 million deaths. According to WHO, the regions most affected by the epidemic were Europe, the Americas, and Southeast Asia, with more than 210 million, 152 million, and 57 million cases, respectively [4]. In addition, the total number of cases and deaths is presented in **Table 1.1** to show the 10 countries that have been hardest hit by the outbreak. The total number of ongoing cases by region between April 2020 and April 2022 is shown in **Figure 1.1**.

Table 1.1: Total number of cases and deaths by country – region until the date of April 2022 [4].

Country Name	WHO Region	Cases - a cumulative total	Deaths - a cumulative total
Global		504,079,039	6,204,155
USA	Americas	79,944,656	981,834
India	South-East Asia	43,047,594	522,006
Brazil	Americas	30,261,088	662,026
France	Europe	27,027,167	141,007
Germany	Europe	23,658,211	133,308
The United Kingdom	Europe	21,863,948	171,878
Russian Federation	Europe	18,101,986	374,141
Republic of Korea	South-East Asia	16,583,220	21,520
Italy	Europe	15,758,002	161,893
Turkey	Europe	15,003,696	98,610

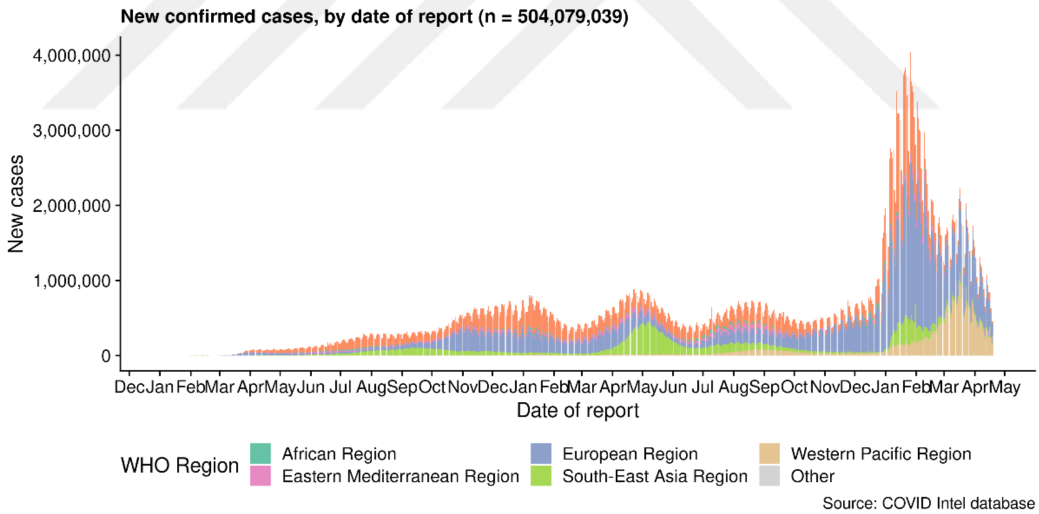

Figure 1.1: Number of ongoing cases by region [4].

Due to the uncertainty of its expansion rate and termination date, the disease, which is considered the world's worst epidemic in the last century, has sparked numerous arguments since the beginning of 2020 [7]. Although the coronavirus disease was not seen as a major threat by society at first, as a result of actions done to lower the infection during the vaccine development process, the epidemic has caused harm to the social development of countries by restricting human freedom in many different aspects, economically, culturally, educationally, and emotionally. Furthermore, as it is well known that the virus spreads easily among people, policymakers have started to seek different solutions constantly [8], [9]. Even though it is impossible to tell when the epidemic will cease, it is evident that after the pandemic crisis is over, various negative economic, social, and health consequences will ensue. Therefore, governments must determine the strategic steps to be taken both during and after the pandemic by listening to the public's voice [10]. Therefore, by emphasizing the importance of post-pandemic information management, researchers identified the main issues that turned into a problem during the epidemic and should be focused on after the outbreak is over [11]:

- Employment and work,
- Education and e-learning,
- E-health and safety,
- E-commerce and consumption,
- Business and economy,
- Entertainment and welfare,
- Environment,
- Equality.

1.1.1. Restrictions and precautions

In order to alleviate the epidemic, which spread rapidly and showed its negative effects all over the world, scientists-researchers emphasized the necessity of the vaccine, therefore, vaccine studies were started to end the pandemic. However, until the vaccine appeared and was applied to the majority of people, governments had applied different restrictions, to protect the public. Different steps taken have been short-lived by causing different negative effects in societies, COVID-19 has shattered the world's daily routine and lifestyle. As a result of this, “what would be the most suitable treatment and prevention method” has been the subject of discussion all over the world [12], [13].

Because it is well known that the virus transmits swiftly amongst people through touch, the first move taken to halt the outbreak from spreading was to quarantine society by establishing curfews. However, concerns about whether the public would get tired of the measures restricting their freedom made policymakers hesitate to start at the beginning of the epidemic to enact the curfew [14]. With the WHO's declaration of the epidemic on March 11, 2020, many countries started to implement curfews, quarantine people, and restrict their social lives forcibly. This situation caused negative feelings in people because of the unpredictability of the current situation, conversely, it helped to keep stable as well as a reduction in the number of cases for a short time. Further, due to the negative consequences of the restrictions, governments started to look for different ways in the vaccine development process and had to stretch the curfews. With the relaxation of the restrictions, the number of cases has begun to rise once more. The number of cases that changed in the countries with the biggest number of instances in the world that changed during and after the curfew, which was generally carried out between March 2020 and April 2020, is shown in **Figure 1.2**.

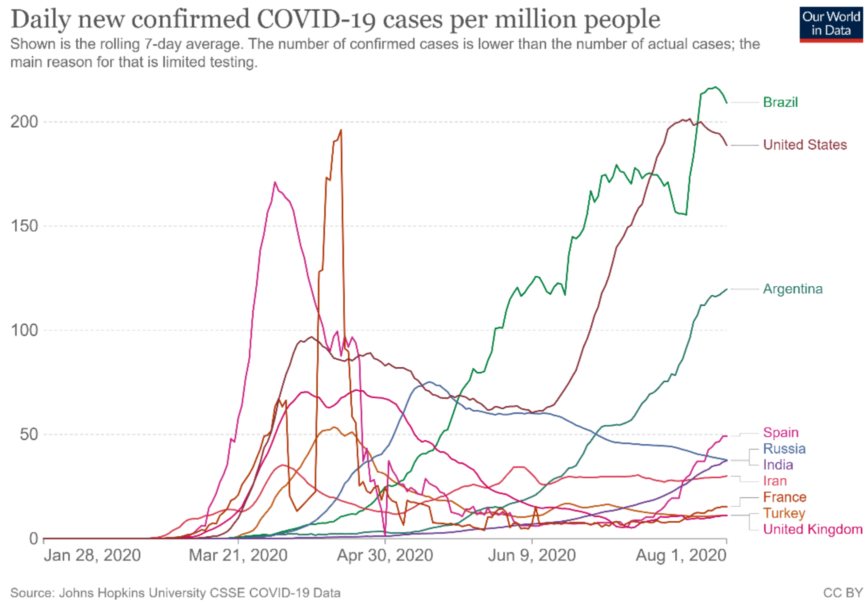

Figure 1.2: The effect of curfews on the number of cases [15].

Different strategies have been tried to be developed to prevent the epidemic by stretching the curfews due to negative consequences. Thus, to prohibit the spread of the virus by people who go out as before, it has been made compulsory to wear masks for everyone in most countries, strict social distance measures have been taken and public awareness-raising information has begun to be shared. With these measures taken, the number of cases decreased again, but it still did not lose its speed of spread. Also, the pandemic has become more dangerous by emerging with different variants in this process.

One of the most serious issues that occurred with the effect of restrictions during the pandemic period was the closure of schools and workplaces. While many schools could not continue their education during the first outbreak of the epidemic, workplaces had to stop their processes. During this period, both students and parents faced many difficulties, their lives were completely restricted. As a result of this circumstance, families' psychology was deeply affected. Later, by establishing online systems, a solution was tried to be created while waiting for the vaccine development, however, the anticipated efficiency was not obtained. Therefore, in the current situation, many schools and workplaces continue their routines in a face-to-face or hybrid way, but to avoid situations where people are so intertwined around the world, the indicator in which bringing processes online is determined in terms of pandemic situation is presented in **Figure 1.3** and **Figure 1.4** by the date of 27 October 2021 that after the start of the pandemic situation.

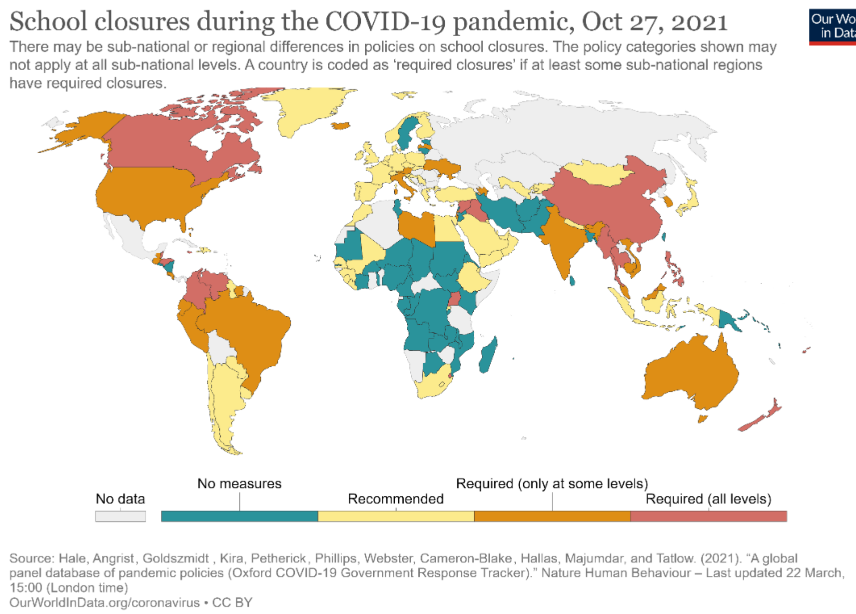

Figure 1.3: School closures status during the pandemic [16].

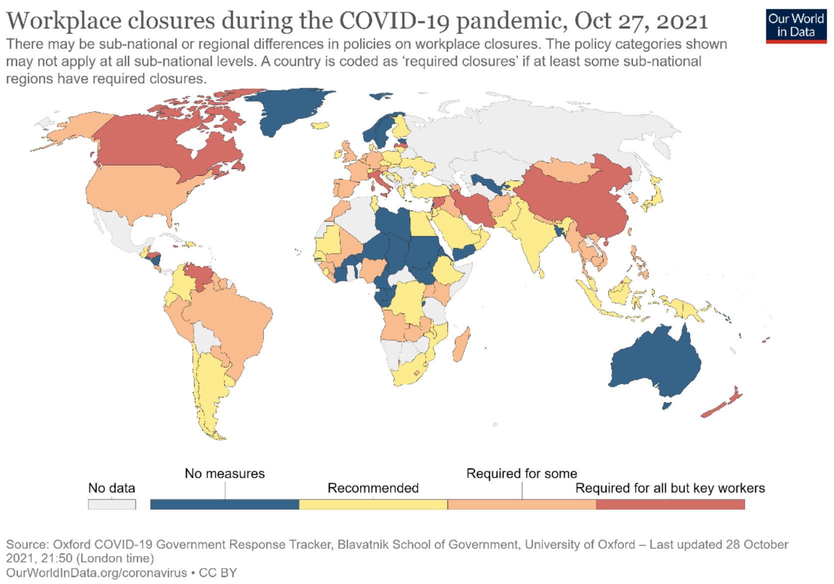

Figure 1.4: Closure status of workplaces during the epidemic [16].

1.1.2. Children, parents, and COVID-19

Due to the spread of the pandemic, efforts have been made to lessen the impact of the virus, consequently, parents and children have experienced interruptions in their daily lives. Although all age groups faced the risk of COVID-19, families with children had to cope with the stress brought on by the loss of psychological and economic support. As a result, the existing situation negatively affected the mental health of the parents, while causing differences in the behavioral movements of the children [17], [18].

Many families around the world have started to experience difficulties such as loss of income, unemployment, care burden, and illness as a result of the sudden outbreak. Parents faced financial difficulties and entered into a process full of anxiety for the health of their unprotected children [19]. In the data created according to 5-year age groups based in 82 countries, the number of people who died as a result of the epidemic (see **Figure 1.5**) and the presentation of the total population with COVID-19 cases in 105 countries (see **Figure 1.6**) are presented, revealing the situation of children compared to adults.

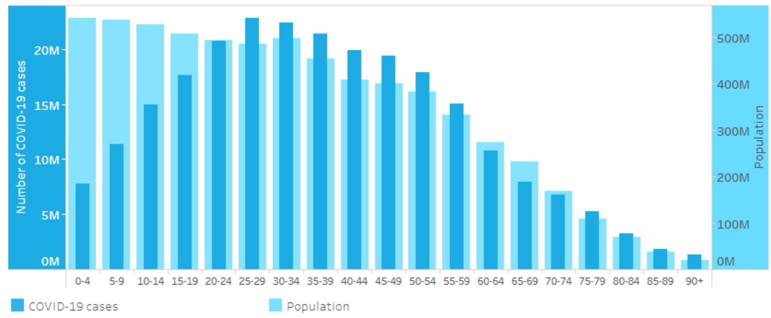

Figure 1.5: The number of deaths caused by the epidemic in the data created according to the 5-year age groups based on 82 countries [20].

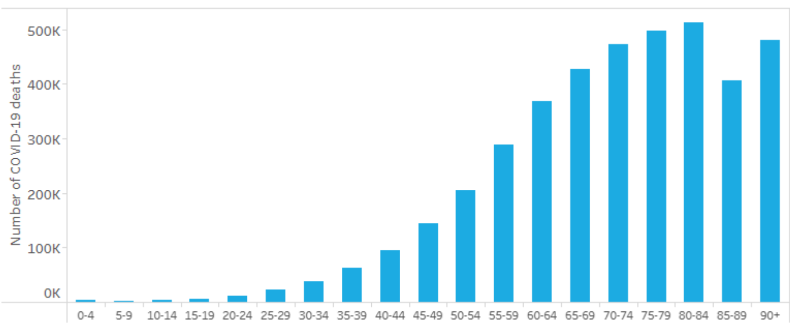

Figure 1.6: Presentation of the total population with COVID-19 cases by 5-year age groups in 105 countries [20].

The number of cases and death rates in children is low compared to adults, but statistics are sufficient to alarm families since children are vulnerable individuals. In addition, when the psychological effect of the pandemic is examined based on age, it is a prominent fact that the negative results in children are more. Children were deprived of their social activities as their lifestyles were affected by situations. Having to keep up with a process out of the order they are accustomed to, they faced changes such as family diseases, social isolation, and online education. As a result, during the epidemic, they began to experience stress, anxiety, and depression [21], [22]. In addition, their personal space has been narrowed due to the intense worry that families have about themselves [23].

Undoubtedly, the situation that most affected families and children in this period was the change of education, which is a measure done to halt the spread of the outbreak. Thus, schools, where children spend most of their time, were closed, and education was stopped until the new system was established. Families, children, and governments have been given great responsibilities, it has been tried to adapt to the new system quickly, online education has started with a rapid transition and steps have been taken to help students attempt to get through this process with the least amount of damage possible. However, there have been many challenges brought by online education [24], [25].

In order to deal with these issues, families shared their concerns about the behavioral changes they see in their children during the process of the pandemic situation, the effects of the new system, how to keep their children safe from the infection, and the side effects of their children about the disease, through social media, as the general public does. They followed the developments related to the epidemic from online platforms. In this way, parents tried to predict the future and tried to protect their children against all negativities by closely observing the world.

1.1.3. Social media and COVID-19

Social media, which is actively used by 4.48 billion people, i.e., 56.8% of the world today, guides users in business, life, and news, also, enables them to follow the agenda by staying up to date. For this reason, it has evolved into a must-have communication tool for the public during the coronavirus. Online platforms have become an invaluable choice for people to share their perceptions transparently, also, to receive and disseminate information. With the rapid spread of COVID-19, social media has begun to reflect the concerns of society [26], [27]. As a result, social media has been actively used by policymakers for the participation and communication of the public in the process, highlighting people's attitudes while taking strategic steps. In this way, to stop the infection from spreading, more solid measures were made.

1.1.3.1. The power of social media platforms in the pandemic process

During the epidemic, while societies were attempting to adjust to the new system with the effect of the restrictions they experienced, they turned to social media and started to share all their worries, concerns, thoughts, and attitudes on social media. In this direction, different platforms have been used for many purposes both by policymakers to understand the perception of the public and by the public to obtain information by observing other people's experiences. During this period, information about the epidemic from all over the world reached everyone most transparently, strategic steps were taken accordingly, and public opinions were shaped according to the direction of the shares. However, it has been observed that social media platforms are also used to misinform societies and to spread deceptive messages that cannot be verified [28].

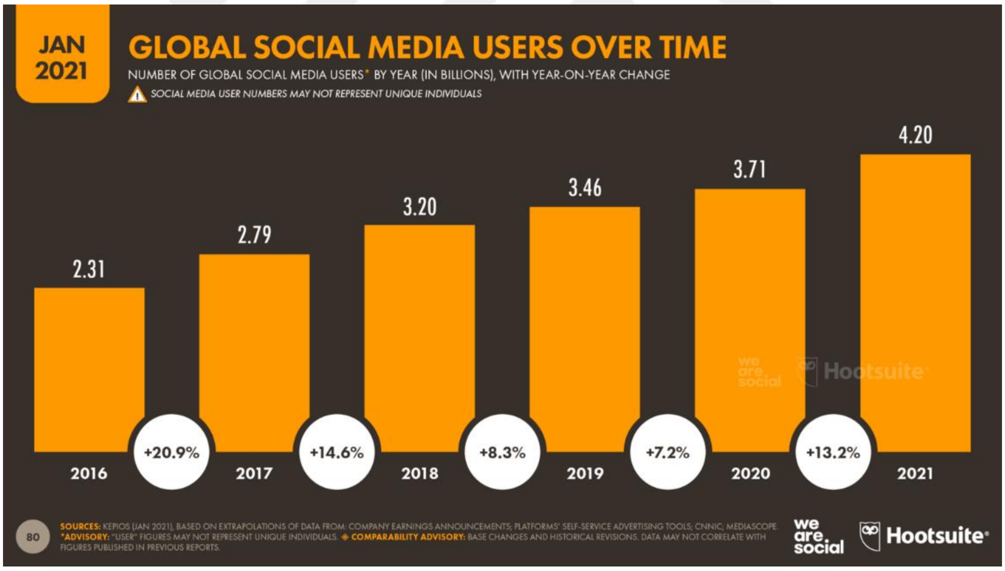

Figure 1.7: The increase in the number of social media users in the last 7 years [29].

The graph showing the increasing number of users in the last 7 years is presented in **Figure 1.7** to show the power of social media platforms. Although many social media platforms were used, during the epidemic, the website that the public turned to was Twitter which had fewer users than the leading online platforms (see **Figure 1.8**). The main reason behind this predicament was that people wanted to follow the latest news about the world and their own countries transparently through a single platform to stay up to date, and also, to observe new developments that emerged during the process.

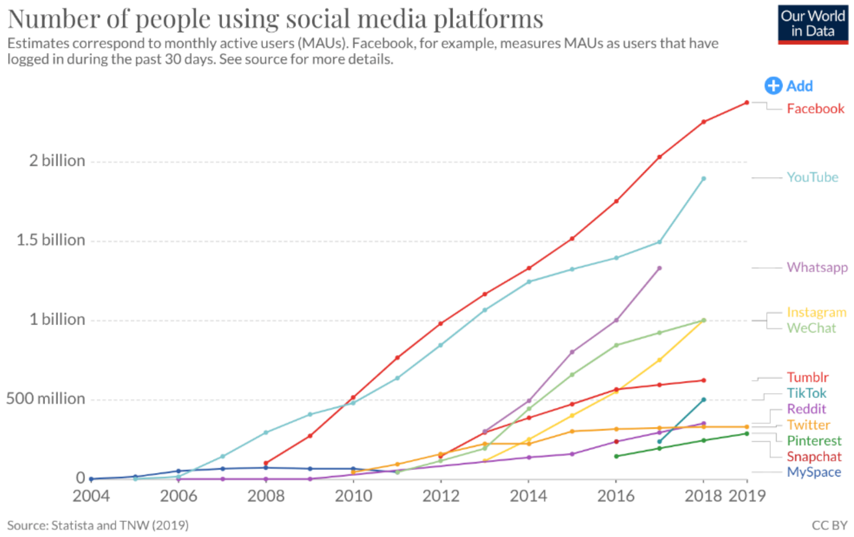

Figure 1.8: User rates of social media platforms [30].

Additionally, since traditional survey methods are time-consuming and expensive, researchers have turned to new techniques to analyze the perceptions of the public. Thus, in order to reach people from all over the world, data sets were created from social media platforms, which have billions of users, and analyzes were carried out based on these data [31].

1.2. COVID-19 Vaccine Development

COVID-19 continues to spread its negative impact on the world and its end is unpredictable for humanity. Most of the strategies introduced to save societies from being infected deprive people of immunity to the virus [32]. For this reason, coronavirus vaccines, which are developed and under development, are seen as a great hope to end the pandemic [33]. It is of great importance for societies to approach these vaccines consciously and to have them vaccinated in full doses to gain herd immunity. Currently, it is known that a total of 11,324,805,837 doses of vaccine have been administered in the world [4]. It is important to mention that to achieve the desired herd immunity, more people need to be vaccinated with a complete vaccine protocol, especially in geographical areas where there are a large number of cases. The graph showing the rate of vaccinations made by the countries that have the highest number of cases in the world and the total vaccination rate in the world is presented in **Figure 1.9**. While it is observed that the rate of vaccination is lower in countries that have a high number of cases compared to other countries, it is thought that the initiatives to increase public awareness against vaccination should be increased by policymakers.

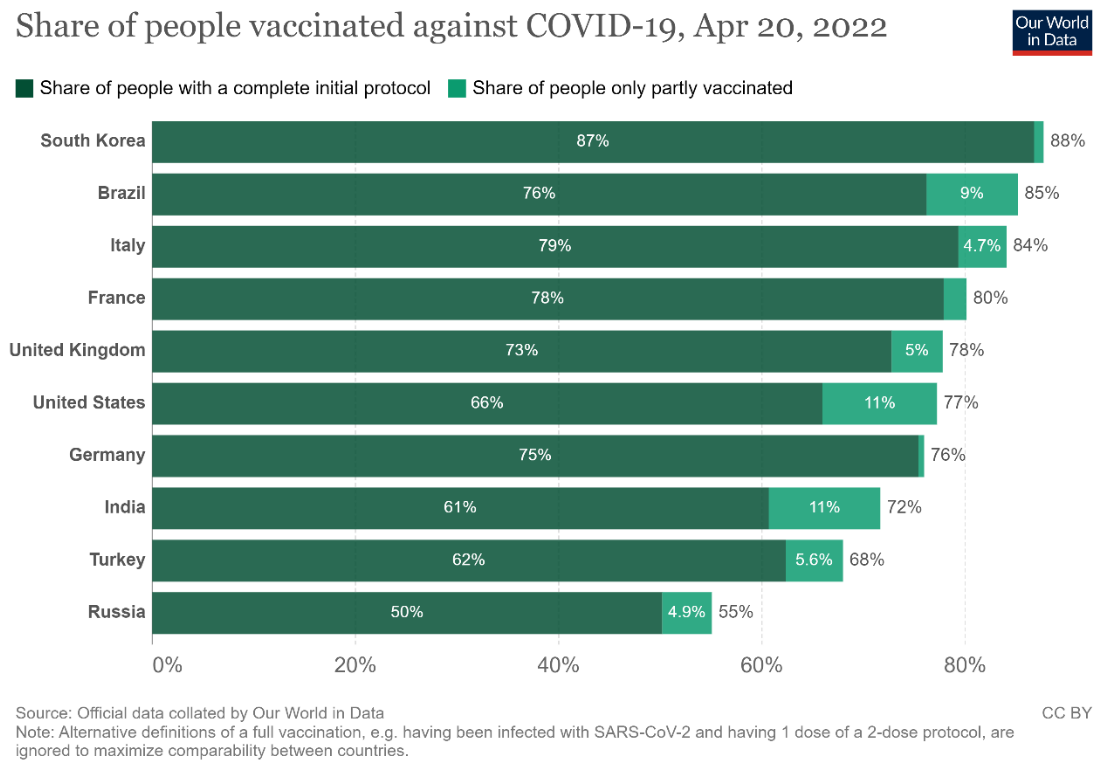

Figure 1.9: Vaccination rates by country and world [34].

With the trigger of the epidemic, researchers have emphasized the need for rapid development of a vaccine that is safe and effective enough to vaccinate an extraordinary number of people to prevent the threat of morbidity/mortality faced by the entire world. Simultaneously, it is thought that more than one effective vaccine should be developed rather than a single approach, due to the global need for vaccines and the wide geographical diversity of the epidemic [35]. Also, since this disease is new to humanity, protective immune responses are poorly understood, so it is not predictable which vaccine strategy will be more successful. In this context, governments and pharmaceutical companies came together and started the development process of many vaccines. In order to be safe from the pandemic and to prevent possible waves, it is expected to provide an immunity rate between 60% and 70% worldwide [36]. Vaccines currently known and accepted by WHO are itemized in the ongoing part of the study [37];

- Pfizer/BioNTech Comirnaty,
- SII/Covishield,

- AstraZeneca/AZD1222,
- Janssen/Ad26.COV 2.S,
- Moderna,
- Sinopharm,
- Sinovac-CoronaVac.

1.2.1. Ongoing debates about vaccines

The pandemic's long-term control is contingent on availability, the development of a preventive vaccine, and a high rate of vaccination [38]. However, since the coronavirus disease is new, how to prevent it is not known exactly, also, its end cannot be predicted. There are ongoing debates that contain certain main themes of the public against existing vaccines and vaccines under development. It is regarded as a fantastic chance for the epidemic to be alleviated by identifying and understanding the negative thoughts arising from hesitation, turning them into positive ones, and encouraging people to be vaccinated.

The "vaccine hesitancy" which is a reluctance to have vaccines, has turned into a problem with the development of COVID-19 vaccines, thus, it has become the most talked-about topic in the process. This hesitation of the people constitutes a major obstacle to the vaccination of the whole society against the rapidly spreading coronavirus [39]. Confidence, peace of mind, risk estimation, collective responsibility, and suitability are decisive factors for vaccine hesitancy [40]. Therefore, reliable sources of information are needed to encourage the population to be vaccinated. Transparent explanations to the public and the development of evidence-based health communication strategies are of great importance in the mitigation process of the pandemic so that the perception of the public is not negatively shaped due to concerns arising from insecurity [38].

Vaccine acceptance is a prominent concept because vaccines developed against COVID-19 disease are seen as a hope to end the epidemic and have a critical role to play in the pandemic's successful control [41]. Accordingly, Sallam [42] conducted a study and examined vaccine acceptance rates through surveys researchers conducted around the world (see **Figure 1.10**). Although the researcher observed that the acceptance of the vaccine is more than fifty percent in general, the researcher stated that understanding the concerns and attitudes about the areas where people are hesitant towards the vaccine is a critical point on the way to gaining herd immunity. Additionally, studies in the literature have consistently emphasized the importance of understanding perceptions toward coronavirus vaccines, which can guide policymakers for future strategies and build trust in vaccines [43].

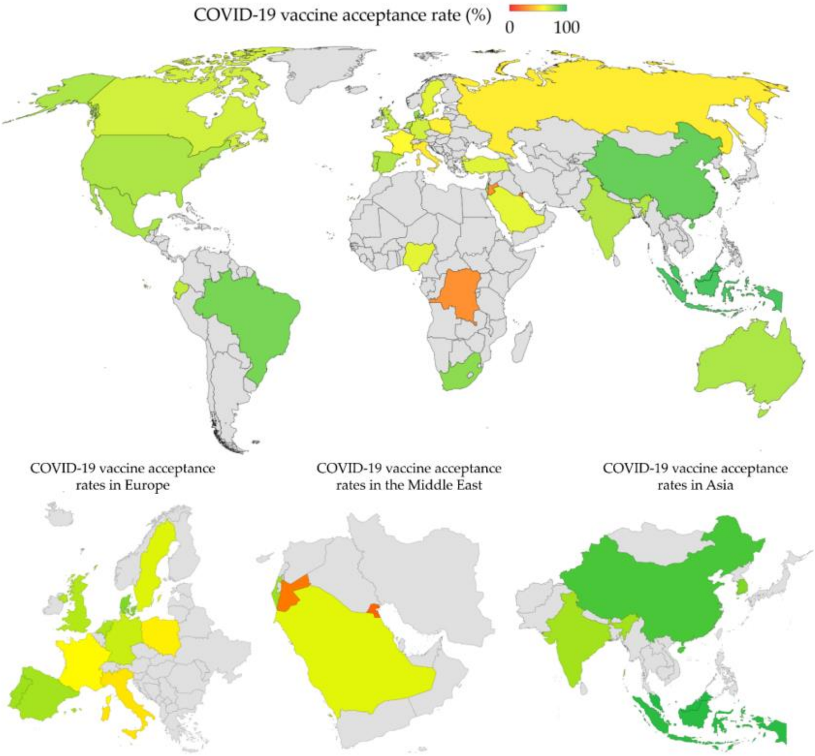

Figure 1.10: Vaccine acceptance rates worldwide [42].

1.2.1.1. Critical elements of effective vaccination

Vaccines, which are thought to be defensive against the pandemic, are of great importance in slowing the course of the epidemic. The hesitant approaches of the people are slow down the process and an obstacle to the acquisition of herd immunity. In contrast, the people who accept to be vaccinated help to slow the spread of the epidemic. However, vaccination rates still have not reached the desired levels. Identifying the main actors that trigger an efficient vaccination process and identifying and solving the points that create a question mark in the public is a critical point for the world to cope with the epidemic. When the different studies on vaccination are evaluated, it has been discovered that certain critical elements in the process and the uncertainties belonging to these elements are discussed to focus on them. In this regard, McKinsey, which covers the topics discussed and presents a summary, has determined the critical elements of an effective vaccination program by working with health experts in the relevant field, while also listing the uncertainties associated with these elements by creating a comprehensive framework [44]. (see **Table 1.2**)

Table 1.2: The critical elements of effective vaccination and their associated uncertainties [44].

Available	Approval of the vaccine and availability in sufficient quantities to reach the population	Technology document-tation and communi-cation	Technology transfer and drug production	Upstream/ downstream sourcing and production public policy	Public policy plan-ning
Administrable	Availability of suitable people to get vaccinated at appropriate places	Population segmenta-tion	Vaccine distribution strategy		
Accessible	Vaccine distribution and storage for use	Order	Logistics, transport, and storage		
Acceptable	Consumers' access to the right information	General communi-cation, messaging, and training	Health workforce training		
Affordable	Affordable vaccination and administra-tion costs	Financing	Repayment strategy		
Accountable	Patients receive a full treatment course and monitor post-launch results	IT infra-structure and interopera-bility	Continuous monitoring and reporting		

1.2.2. Parents' perspectives on vaccines for their children

With the reopening of schools on the agenda, the debate about the vaccinations to be applied to children has become more intense, since children will be completely unprotected, also, measures such as masks and distance are not sufficiently protective. The easing of the course of the epidemic is a critical point for children, whose social lives are completely restricted due to the isolation, pandemic situation, and online education, to continue their lives as normal [45]. Vaccines are regarded to be the most essential strategy to accomplish this situation [46], [47]. However, there is an existing worry and distrust among parents about getting their children vaccinated. Considering all aspects, families' attitudes towards vaccination of their children should be evaluated, and strategic steps should be taken in this direction.

During the epidemic, it was determined that the willingness of parents to protect their children from COVID-19, who started to return to their normal lives with certain precautions, triggered the acceptance of vaccines. It was also discovered that vaccination apprehension arose as a result of concerns about the vaccine's negative effects and a lack of public information about the vaccine's effectiveness [47]. Considering the studies conducted in the literature with specific samples based on questionnaires, it is critical in the current situation to share the development process of coronavirus vaccines transparently with the whole society.

1.2.2.1. Factors affecting parents' perceptions of vaccination

Some factors that determine the acceptance-rejection rates of families to vaccinate their children come to the fore. The socio-cultural structure of the society, and the geographical regions in which they continue their lives shape their perspectives on vaccines, and heavily affect their attitudes, ideas, and perceptions. In general, the fact that vaccine safety and efficacy are not fully known creates an uncertain situation in the process, this uncertainty draws the perspective of families to the vaccination status of children in a negative direction. The domestic or foreign production of vaccines is one of the factors that influence perceptions, parents prefer domestic vaccines compared to foreign vaccines and have more confidence in domestic vaccines [47]. Another factor is the clear observation of the positive effect of vaccines on humans. It is thought that the course of the COVID-19 vaccines in easing the spread of the epidemic will affect the vaccination acceptance rates of parents who want to protect their children against the virus [48].

As a result, it is crucial for the whole society to fully understand the perceptions of families against vaccination. To the best of our knowledge, the outcomes of the investigations in the literature are gathered around certain points, since they were studied with small samples, in certain populations, and the questions directed to the parents were determined by the researchers. For the process to progress successfully, it is thought that models should be developed that will deal with the world in general, produce more comprehensive results, provide the emergence of different points, and consider the changes in emotions.

The organization of thesis is organized as follows: in the second chapter, the theoretical part that forms the basis of the thesis will be discussed and a comprehensive summary of the literature will be given by examining the studies in this field. In the third chapter, an experimental part will be created, the framework of the methodology used will be drawn, the techniques, programs and codes used will be explained. In the fourth chapter, the outputs of the application will be shared, and the results of the thesis will be presented in detail by discussing the findings. In the last chapter, a conclusion will be made about the study and suggestions will be made to guide future studies.

CHAPTER 2

2. THEORETICAL PART

2.1. Literature Review

Researchers have turned their focus to COVID-19 which has become the most important global problem of recent times and brought many studies to the literature due to public concerns about this epidemic. The authors aimed to determine and analyze the echo of the COVID-19 around the world, the public attitude towards the coronavirus vaccine, the perception of the parents about their children in this period, and the attitude of the parents towards a possible vaccine to be given to their children. Thus, they carried out their research by adopting many different data sources, methods, and perspectives.

The purpose of the literature review was to observe the debates, mood changes, and attitudes caused by the pandemic in the public, and also, to determine the public reaction to a possible or existing COVID-19 vaccine. It also has the goal of determining the popular methods that the researchers included in their studies within the relevant subject's breadth and examining the tools that support the application while performing the application. In this direction, while highlighting the importance of the subject, the studies are classified according to the subjects and methods they are based on, to identify the missing aspects in the literature.

With the help of keywords related to epidemic and coronavirus vaccine, Google Academic, Web of Science, Istanbul Medipol University Library-Electronic Information Resources were scanned. As the articles were reviewed, the scope was narrowed down with keywords based on different perspectives and methods about the epidemic for the selection of appropriate literature. Within the scope of this thesis, more than 80 studies based on the relevant subject were analyzed and investigated. Because of the subject's popularity, the literature study was conducted between 2020 and 2021. As a result, the articles published within this timeframe were included in the study.

It was established that studies conducted in the early days of the pandemic were focused directly on the epidemic and also discussed policymakers' actions around the world or in a specific country. On contrary, it was observed that the scope of the subject was expanded in the future, and different perspectives were included in the studies conducted. Simultaneously, it has been determined that many researchers are considering the COVID-19 vaccine, investigating the public perception in case of a possible vaccine, or guiding policymakers about the existing vaccine. In this context, it has been discovered that recent studies have been conducted with the assistance of social media platforms, such as government, health, and individual social media accounts, as well as leading news sources. It has been sighted that they often make use of NLP techniques. As a result, it has been observed that researchers have benefited from topic modeling techniques and sentiment analysis methods, with the support of Twitter, to highlight the main issues related to the pandemic, the sub-topics covered by the main topics, to analyze the mood changes of the public and to identify critical situations in taking steps for the future.

2.1.1. COVID-19 echo in the public

COVID-19 has completely changed the lifestyle of people in different regions and countries; thus, it has become a very important global problem. Although it has been a long time since the emergence of the pandemic situation, the cases and deaths due to the disease continue. This situation leaves a deep impact on different aspects such as the psychological state of the people and the global economy. In this context, the researchers, who emphasized the importance of understanding the repercussion of COVID-19 in the public, focused on this field and tried to determine the developing perspective of the people towards the pandemic and the attitudes towards the strategies of policymakers.

Singh et al. [49] aimed to develop a perspective on the pandemic by addressing social media platforms, especially Twitter, also, focusing on the number of conversations about COVID-19, the themes of the topics, the thoughts shared about the epidemic, and how much these shares were shared by other users. As a consequence of the research, it was discovered that there is a strong link between new instances and information spreading on social media. There are discussions about false information and low-quality posts, although they are less than in other crises.

Using the survey web link on Instagram, Twitter, and WhatsApp, Ali et al. [50] aimed to analyze the behavior, knowledge, and perceptions of the public regarding the epidemic to identify deficiencies in key parts of public information. As a consequence of the investigation, while it was observed that the well-informed population applied the correct protective measures, it was determined that the relevant group had basic knowledge deficiencies about vaccination against COVID-19.

For future reference, Cuiyan et al. [51] conducted a study in China to evaluate the psychological effects of the epidemic on the population, including depression, anxiety, and stress levels. In this context, it was discovered that the majority of respondents ranked the psychological impact as moderate to severe, while a minority rated the anxiety as moderate to severe, as a consequence of the online surveys conducted.

In a study conducted in the UK, USA, and Singapore, Raamkumar et al. [52] tried to determine the reactions to the pandemic-related relief efforts of public health officials by receiving support from Facebook. Data were drawn from the profiles of public health officials for three countries, and it was discovered as a consequence of the investigation that the scope of use of the profiles differed, while it was observed that social media analysis could provide information to the public about the communication strategies of health organizations during the pandemic.

With the help of online surveys conducted in the USA, Fitzpatrick K et al. [53] aimed to analyze the association of fear of epidemics with mental health outcomes and social security deficits. It was found that the majority of those who took part in the study were afraid of COVID-19, the participants showed depressed mood, and the participants, especially those who were socially vulnerable, faced mental health problems in addition to the feeling of fear.

In Korea, Park et al. [54] aimed to analyze the news sharing behaviors and information transfer networks on Twitter about the epidemic. In order to identify the main problems of the epidemic, the authors investigated the problems comparatively with the help of network analysis. It's been established that the word "coronavirus" is used the most as a keyword in social media, also, the majority of news shares draw the reader's attention to the positive behaviors of individuals, but news with a medical basis is rare. Although it is rare, it has been observed that the potential for dissemination of medical news in public is higher than in other news.

Table 2.1: The data source used by the studies.

Author	Data source	Purpose	Result
Singh et al. [49]	Twitter and other social media platforms	Presenting a perspective on the pandemic	There was shown to be a link between the new cases and the discussions on the platforms.
Ali et al. [50]	Online surveys posted on Instagram, Twitter, and WhatsApp platforms	Determining public attitudes towards the outbreak	It's been noticed that the well-informed population acts consciously but lacks knowledge about vaccines.
Cuiyan et al. [51]	Online surveys	To determine the psychological state of the people during the epidemic	The majority of the individuals were found to be moderate to severely affect psychologically.
Raamkumar et al. [52]	Facebook	Observing the reactions to the relief efforts carried out during the pandemic	It has been put forward that those emotional changes fluctuate according to the shares and their scopes differ.
Fitzpatrick K et al. [53]	Online surveys	Evaluating the psychological consequences of pandemic anxiety and its relationship with social security vulnerabilities	It has been determined that the feeling of fear against the epidemic is generally dominant.
Park et al. [54]	Twitter	Examining news sharing behaviors and information transfer networks	While "coronavirus" is the most used keyword as the keyword, it has been analyzed that news posts rarely have a medical basis.

In order to analyze the public's thoughts and discourses about the epidemic, support is often received from online environments. In addition, the traditional method of surveys was carried out from online platforms to find out what the general public thinks, thus, the opinions of the public are taken into consideration transparently (see **Table 2.1**). It was as a result of the investigation that a common point reached by the authors was that the people were more worried during the epidemic increase and that they looked more hopeful for the future outside these periods. During the pandemic process, the predominance of news that does not reflect the truth and low-quality shares was observed. The distribution of news and the accuracy or falsity of information presented to the public was considered. It has been discovered that using social media, particularly by politicians, health groups, and states, is more effective in making public remarks and in informing the public about the epidemic.

2.1.1.1. Parents and children in the COVID-19

As the coronavirus epidemic spread around the world, governments have taken some protective measures to prevent the further progression of the outbreak. As a consequence of the strategic steps taken by policymakers, some obstacles have emerged in the lives of parents and children. In this context, researchers have set themselves the goal of examining the troubles experienced by parents during the pandemic and the effects of these troubles on children, as well as analyzing the concerns of families for their children.

Until the beginning of June 2020, Patrick et al. [17] aimed to analyze how the pandemic in the USA and its efforts to alleviate the pandemic changed the emotional-physical state of families. As a consequence of the study's questionnaires, it was discovered that the pandemic had a significant impact on both parents and children. In this direction, it was emphasized that policymakers should take additional measures to mitigate the effects of the epidemic that may occur in the economic and health fields, while it was suggested that the needs of parents who have children should also be considered.

Zhen [55] conducted a research study to understand parents' perceptions of the consequences for their children and families during the epidemic. By wrestling with 6 different parents from different educational and social backgrounds, it was determined that younger children from families who were aware of the effects of the epidemic, although not long-term, showed some results in terms of education, physical/mental health, and children exhibited some reactionary behaviors in general.

Yeasmin et al. [56] aimed to analyze the child mental health consequences of the epidemic in Bangladesh during the curfew period. Between April 25, 2020, and May 9, 2020, 384 parents with at least one child aged 5-15 were a part of the research. It was discovered that the majority of the children's mental health deteriorated during the quarantine period, they experienced depression, anxiety, and sleep disorders during this process.

In the pandemic, serious consequences have emerged for families, especially those that have children. Consequently, the direct effect of the epidemic on their psychology has been observed. As a result of surveys or interviews with parents, it was discovered that the COVID-19 showed some results in terms of education and physical/mental health in children. It was observed that during the closure period, children faced depression, sleep disorders, and anxiety problems, therefore their mental health deteriorated. With these observations, it was emphasized to the governments that steps should be taken to mitigate the epidemic and that the development of strategies to relieve people would positively affect the course of the epidemic.

2.1.1.2. COVID-19 vaccine perception in the world

It has been stated that developing vaccines early in the COVID-19 pandemic is crucial for preventing the pandemic, although it is projected that 67 percent of the population must be inoculated for the vaccine to be effective in curbing the disease's spread [57]. Coronavirus vaccines have attracted great interest from the public, subsequently, they have led to many discussions that have brought about concerns and hesitations. Based on the current situation, researchers emphasized the importance of understanding the public's views towards the vaccine while the epidemic was still ongoing and tried to analyze their perception. In this way, they aimed to guide policymakers by examining the problems and concerns for increasing vaccine intention and shed light on the debates about vaccines.

Using the theoretical framework of the Elaboration Likelihood Model, Extended Parallel Process Model, and Social Judgment Theory, Scannell et al. [58] examined different vaccine sensitivities as pro-vaccine, anti-vaccine, and neutral. They wanted to figure out how different persuasive strategies were employed in COVID-19 vaccination tweets. By analyzing the Twitter data set consisting of 1,000 rows of data, the researchers found that the tweets posted as anti-vaccine centered on topics such as election categories, political/conspiracy theories, and security, generally used humor, also, sarcasm, famous figures, anecdotal stories as a persuasion technique. On the other hand, they observed that pro-vaccine-related tweets' contents have knowledge, participation, and famous figures as tactics of persuasion. As a consequence of the investigation, a response framework called the Health Information Persuasion Survey was proposed to address anti-vaccine and misinformation.

Lazarus et al. [59] conducted a survey covering 19 countries, including 13,426 people, to ascertain the factors that influence the adoption of the COVID-19 vaccine with human trials in June 2020 and try to estimate the potential acceptance rates. As a result, it was discovered that the part of the participants who stated that they trusted more government-sourced information was more likely to accept the vaccine and to follow their employer's recommendations on vaccination.

In order to comprehensively assess the vaccine hesitancy of the adult population in the USA, Khubchandani et al. [60] took help from social media to get ideas from the general population and conducted online surveys in this context. With the multiple regression analyses as a consequence of the questionnaires, the public hesitations about the vaccine were observed in terms of education, gender, income, having children at home, political views, employment, and fear of being infected in the future.

In a randomized controlled trial in the USA and UK, Loomba et al. [61] aimed to analyze how exposure to false information about vaccines shared on social media impacts vaccine acceptance. As a result, the importance of false information that seemed scientific led to a decrease in vaccine acceptance, fewer people in both countries wanted to be vaccinated than needed, and misinformation circulating on social media, in general, led to a decrease in vaccine intention.

In France, Ward et al. [62] aimed to determine attitudes towards a possible vaccine through online surveys in April 2020, by taking a sample aged 18 years and over in quarantine. While it was determined that a quarter of the respondents were not interested in a possible vaccine, it was observed that the existing attitudes towards the vaccine were strongly related to the political system and partisanship.

Based on metropolises from five countries, Hou et al. [63] aimed to analyze public participation, hesitation, and confidence in vaccination. They examined vaccine-related public communication with the support of a framework based on data obtained from social media and the model developed by WHO for vaccine hesitancy. In New York and London, there was a finding that there was distrust of the vaccine, the government, and experts, while there were concerns about the supply, as well as production of the vaccine in Mumbai, Sao Paulo, and Beijing. As a result, the importance of vaccine hesitancy is common worldwide and negative posts get more interaction.

Murphy et al. [43] carried out a study with the help of surveys in Ireland and the UK, emphasizing that identifying vaccine hesitancy in different samples could be beneficial for future public health messages. The presence of vaccine resistance was found to be

35% for Ireland and 31% for the UK. It was observed that those who approached the vaccine hesitantly in both samples were less likely to receive information from authorized sources about the epidemic and were distrustful of the information they received.

In May 2020, Reiter et al. [64] conducted an online survey of persons aged 18 and up in the United States to assess public perception of the vaccination. It was determined that 69% of the participants were positive about getting vaccinated. Researchers emphasized that vaccine acceptability should be monitored by the public during the vaccine development process.

In the USA timeframe from April 16, 2020, to April 20, 2020, Fisher et al. [2] was conducted a cross-sectional study with a representative sample of adults, aiming to determine vaccination hesitancy and intention to be vaccinated, as well as, to observe the reasons behind the thoughts. As a result, it was observed that 3 out of 10 respondents were unsure about the vaccine, and 1 out of 10 participants did not consider getting vaccinated. Consequently, it was emphasized that multi-faceted and targeted strategies would be needed to increase the intention to be vaccinated in a possible vaccine.

Through surveys conducted with the support of an online platform, in the USA in May 2020, Malik et al. [33] aimed to determine the risk perceptions about the epidemic and the acceptance of the vaccine. As a result, although it was determined that the vaccine was accepted at a rate of 67%, it was determined that there were demographic changes in the intention to be vaccinated.

In order to analyze the perception of vaccines, researchers frequently preferred online questionnaires and aimed to determine the existing attitudes by asking specific questions about the vaccine. The most striking point in the studies conducted is that people approach the vaccine with hesitation and have negative feelings, while it has been observed that the current uncertainties cause this. The main reasons for vaccine rejection are the public's concerns about vaccine production-supply processes and the vaccine's lack of confidence due to false information circulating on social media. Furthermore, it has been stated that attitudes towards vaccines have a serious relationship with political views, and it has been asserted that the reasons behind the opposition to vaccines stem from political views. Although the minority of studies found vaccine acceptance to be high, they stated that they encountered great geographical differences. The truth has been emphasized that the public has a high level of trust in scientific information and the government. It has been suggested that targeted and multifaceted strategies will be needed to increase vaccine acceptance for the sake of gaining immunity in the public.

2.1.1.3. Vaccination hesitancy of parents for their children

In the current literature, the introduction of a positive and safe COVID-19 vaccine has been shown as a way to support the end of the disease. However, the issue of vaccination in children remained unclear and caused hesitation in families with children. Despite the recognition that vaccination is an effective and safe approach to preventing communicable diseases, many questions have arisen in the public about vaccination, which has led to heated debate among parents, who have difficulties in determining which condition is safer for their child. These questions include uncertainty about the efficacy of the vaccine among children under the age of 18, vaccine skepticism, and vaccine skepticism regarding the uncertainty of side effects [65]. With the emergence of the public's hesitations against the coronavirus vaccine, researchers aimed to analyze the intention of parents to vaccinate their children against a possible or existing vaccine and to determine their thoughts on the subject to provide insight into the future.

Alfieri et al. [57] aimed to compare the hesitations against the COVID-19 vaccine that could happen in the future by considering children from different sociodemographic groups, while also aiming to understand how parents obtained information about the pandemic. Surveys were taken from 1,702 parents and analyzes were carried out with 1,425 of them. As a result, the greatest rate of hesitation was found in the demographic groups most severely affected by the epidemic. It has been suggested to these groups that outreach efforts should be made from reliable sources of information for COVID-19 promotion.

Yigit et al. [47] aimed to predict the frequency of vaccine refusal by identifying the reasons behind the hesitation and rejection of COVID-19 vaccines in parents. By asking 428 parents about their families' sociodemographic characteristics, their views on COVID-19 vaccines, and their reasons for refusing the vaccine, it was discovered that the majority of the participants were wary of immunizations. As a result, the researchers commented that the uptake of vaccines may increase with a greater preference for domestic vaccines.

Considering that the successful application of the vaccine in the USA is due to reasons related to the possibility of the parents and their children being vaccinated, Halvorson et al. [66] conducted a national household survey in the time frame of 5-10 June 2020 with 1,008 samples, in which families were asked about their probability of vaccinating themselves and their children. As a consequence of the investigation, it was discovered that only 60% of the parents had a positive perspective on vaccinating themselves or their children. It is thought that in this direction analyzing the existing hesitations can create herd immunity against the epidemic in the USA.

In the UK, in families with children 18 months of age or younger, by surveying 1,252 parents and interviewing 19 respondents, Bell et al. [48] aimed to measure parents' perceptions of a possible future COVID-19 vaccine, using multi-method approaches including semi-structured interviews, as well as, online questionnaires. In conclusion, it was emphasized that information about the COVID-19 vaccine should be communicated publicly.

To investigate the acceptability of the vaccine by parents in China, Zhang et al. [46] conducted online surveys on a sample of 1,052 parents, and it was found that parents were keen on the COVID-19 vaccine for their children. Researchers emphasized the importance of transparency in vaccine safety testing, also, as the vaccine development process, and stated that false information should be discovered, as well as, resolved on time.

Goldman et al. [67] aimed to conduct an analysis of caregivers around the world about their intention to vaccinate their children in the event of a possible vaccination. It was determined that most of the caregivers participating in the cross-sectional survey were willing to vaccinate their children against COVID-19, but different factors played a role in the vaccination intention. While it is recommended for policymakers to provide the public with evidence of efficacy, as well as safety for a possible vaccine, it is emphasized that barriers to vaccine acceptability can be removed by noting the risks of effects on children and educating caregivers about vaccination.

Skjefte et al. [68] conducted an online survey among mothers of children younger than 18 years old and pregnant women, aiming to monitor vaccine acceptance by country. As a result, it was discovered that the Philippines, India, and Latin America had the highest vaccine acceptability, while the United States, Russia, and Australia had the lowest. The study found that one of the most important elements in vaccine adoption is belief in vaccine efficacy and safety.

Yılmaz and Sahin [69] aimed to determine and analyze a parent's perspective on the vaccine developed against the epidemic. The researchers, who conducted a cross-sectional study with an online questionnaire with the participation of 1,305 parents, observed that the willingness of parents to vaccinate their children was low.

By performing content analysis, Jenkins and Moreno [70] aimed to analyze the shares of parents on social media about getting their children vaccinated. As a consequence of the study, it was discovered that parents chose non-medical solutions, and this trend has grown due to social media's influence.

Table 2.2: The data source used by the studies and the sample size they deal with.

Author	**Data source**	**Considered sample**	**Purpose**	**Result**
Zhang et al. [46]	Surveys	1,502 parents	Observing the acceptability of vaccines	It has been commented that parents are keen on vaccines for their children.
Yigit et al. [47]	Surveys	428 parents	Uncovering the causes of attitudes towards vaccines	It's been proposed that there is greater reliance on domestic vaccines.
Halvorson et al. [66]	Surveys	1,008 parents	To reveal the perspective of families on vaccination	It was determined that 60% of the parents were positive about vaccination.
Bell et al. [48]	Meeting/ Surveys	1,252 parents	Identifying attitudes towards vaccination	It has been suggested to policymakers that they provide transparent vaccination information.
Alfieri et al. [57]	Surveys	1,425 parents	Comparing hesitations against vaccine	It was determined that the greatest hesitation occurred in the regions most affected by the epidemic.
Goldman et al. [67]	Surveys	1,005 caregivers	Revealing intention to vaccinate	It has been suggested that carers are generally willing to vaccinate.
Skjefte et al. [68]	Surveys	17,871 participants	Detecting vaccine acceptance	Efficacy and safety of vaccines have been demonstrated to play a significant effect in vaccine uptake.

Yılmaz and Sahin [69]	Surveys	1,305 parents	Identifying vaccine thoughts and wishes	It has been commented that the willingness of families to have their children vaccinated is low.
Jenkins and Moreno [70]	Social media	244 comments from nine blogs	Examining the parents' shares about the vaccine	It has been observed that parents prefer non-medical interventions.

As a consequence of the research, it was discovered that parents have a high rate of hesitation about getting their children vaccinated. It was emphasized that transparency is important in the vaccination process and that a clear introduction of the vaccine should be made to the public. It has previously been reported that it is very important for herd immunity to eliminate the question marks of parents who are worried about their children and to evolve their ideas positively. Most of the researchers preferred the questionnaire method to examine this issue and asked questions to families about getting a vaccine for their children and themselves or only their children vaccinated (see **Table 2.2**).

When looking at existing studies, it was shown that relatively few use major social media platforms other than blogs to profit from machine learning approaches, as a result, the lack of this situation in the literature has been observed. It is thought that conducting a study on the hesitancy of vaccination in children on Twitter, where the public shares their feelings in the most transparent way, will bring clarity to the subject and bring the issue of children and vaccination to the fore with sharp lines throughout the world. Furthermore, it was discovered that the studies carried out covered a very short period, it was thought that a wider range should be observed for more efficient results, and it would be advantageous to expand the date range to observe the effect of perspectives such as distance education on the idea of the vaccine.

2.1.2. Natural language processing (NLP) studies in the COVID-19

NLP is a field that tries to process, understand and interpret human language. It provides methods that help transform the text into a structured representation and enables artificial intelligence to make sense of it by taking human language as input. In short, they are techniques that enable automatic extraction and identification of information from texts [71], [72]. In this context, to analyze the attitudes of the public against the pandemic, researchers have benefited from NLP techniques and examined the main themes and mood changes of public perception within the historical frameworks they determined. The studies are summarized in terms of the techniques they used and presented in **Table 2.3**. In addition, other studies are included in the continuation of the chapter, and articles dealing with different NLP methods are examined.

Table 2.3: NLP studies on COVID-19.

Author	**The algorithm used**	**Purpose**	**Result**
Masum et al. [73]	Query expansion Data pre-processing Vector space models	Developing an answering framework to automatically analyze thousands of studies to generate long text answers that can answer medical questions	A viable knowledge extraction framework has been presented by processing the used articles with selected techniques.
Klein et al. [74]	Deep neural network classifier based on BERT model	Developing a Twitter-based NLP data pipeline that can self-process from publicly generated social media data to detect potential cases in the USA	In order to identify tweets that automatically revealed potential cases, the automated hotline was activated, resulting in the analysis of 13,714 tweets in the USA.
Ayoub et al. [75]	SHAP DistilBERT	Coping with misinformation during the outbreak	It has been discovered that the likelihood of trusting information about the pandemic is significantly higher in posts containing text&SHAP explanation and text&SHAP explanation&source&evidence.
Bose et al. [76]	LSA	Extracting keywords based on demographic, social, epidemiological, economic, psychological, medical, and clinical perspectives of the pandemic with the support of NLP algorithms, based on the literature	10 main themes have been identified that are COVID-19 response, flu response, COVID-19 testing, COVID-19 management strategies, influenza response, general respiratory syndrome and coronavirus, clinical trial, Ebola response, COVID-19 hospitalization, as well as, general hospitalization, including the progression of the epidemic and help informs a general range of insights into the course of scientific studies.

Tang et al. [77]	BERT	Analyzing the tweets about the epidemic posted by health organizations in Texas, identifying the characteristics of the content, and how these posts can predict the level of responsibility of the public	It has been observed that tweets about the steps taken and the epidemic are more likely to be shared among the people, while tweets about community functions and actions are more likely to be liked.

Sadman et al. [78] collected 1,050 news reports about the pandemic from two different countries and used NLP techniques to obtain information about the number of cases, trending topics, emotions, and the virus. In their study, the authors analyzed how the COVID-19 virus spreads, as well as, how a developing country and a developed country were affected by the pandemic.

L. Li et al. [6] received support from NLP techniques by processing the data pulled from the Weibo platform with the help of supervised learning to classify the available information about the epidemic into different types of situation information. Consequently, status information that emerged based on data were attention and advice, notifications, or measures to be taken, food or service donations, emotional support, help-seeking, doubt and criticism, counter-rumors, and non-situational information. For each species, some features were identified to predict retransmitted quantities.

Samuel et al. [79] aimed to determine the public perception of COVID-19, with the help of sentiment analysis packages of the R program, by analyzing tweets about the pandemic. It was discovered as a result of the application that the feeling of fear in the public increased over time as the course of the epidemic increased in the USA, while the relevant opportunities, limitations, results, and related methods were outlined.

Sun et al. [80] aimed to present an epidemiology model with the support of NLP tools using the action monitoring reports of COVID-19 patients in China. The researchers observed that the developed model would be very useful in putting forward control policies in the so-called "new normal" conditions. Also, they determined that restaurants carry a greater risk than other issues, including places where the population is concentrated.

Osakwe et al. [81] aimed to identify the main issues by using textual data to determine the reaction and concern of the public during the pandemic, with the study they carried out between the dates 10 May 2020 and 24 May 2020. With the support of 7,301 tweets, analysis of the textual data revealed six main topics: surveillance, prevention, treatments, testing and treatment, symptoms and contagion, fear, and financial loss.

Different NLP techniques are used by researchers to evaluate public opinions with the support of social media platforms, to determine the accuracy of information shared by news sources or websites, to determine how much the public trusts the news, information, or ideas that are circulating, and how such articles affect people has been observed. Unlike supervised learning, it has been determined that unsupervised learning, in which the outcomes are unclear and an unlabeled data set can be processed, is more useful for examining large data sets in a broader context.

Topic modeling techniques have come to the fore, as researchers who perform unsupervised learning practice mostly use the. Furthermore, it was discovered that the authors made extensive use of sentiment analysis techniques to further strengthen their research and determine the perception of public opinion, by this means, they had the chance to understand, also, interpret the public's feelings about the pandemic. In addition, it has been concluded that following social media movements can be useful for understanding the public concerns constantly changing during the pandemic period.

2.1.2.1. Topic modeling and COVID-19

Topic modeling, also known as probabilistic clustering, is thought to be one of the most effective techniques for classifying and clustering textual data [82]. Researchers prefer to use topic modeling techniques that make use of unsupervised learning to reveal hidden thematic structures from the data sets they have. A topic model created can introduce a range of interpretable topics and assess the power of each document to showcase those topics [65]. In this context, the authors preferred topic modeling techniques instead of traditional methods to examine public attitudes about COVID-19, reveal the main issues, and analyze the current situation, thus, they produced more comprehensive studies with more data, obtained robust, as well as real results. While scanning the literature, it was observed that some researchers used Twitter as a data source, and some authors analyzed their data by taking their data from different web pages, as a result, these studies were summarized in **Tables 2.4** and **2.5**, respectively.

Table 2.4: Studies using Twitter as a data source in topic modeling studies.

Author	**Time range**	**Number of tweets**	**Purpose**	**Result**
Lyu and Luli [83]	11 March 2020- 14 August 2020	290,764	Identifying themes that related CDC	It was discovered that the most talked about subject are deaths due to the epidemic. 4 main themes emerged: knowing the virus, policy actions, and general opinion on reliability.
Agarwal et al. [84]	1 January 2020- 30 April 2020	866,527	Developing a system for analyzing tweets about the COVID-19	Eight main themes were identified: pandemic impact &reopening, government response, health workers &officials, federal aid &quarantine, the origin of the novel coronavirus, people's thoughts, case statistics, quarantine and stay at home.
Doogan et al. [85]	January 1, 2020- April 30, 2020	777,869	To identify the posts about non-drug interventions by considering six different countries	New Zealand is warier of non-drug interventions, while the US, on the contrary, has received less attention. It was discovered that 4 main issues related to non-drug intervention and implementation strategies, inconsistent information, and timeliness of implementation affect public commitment.

Table 2.5: Studies using web pages as the data source in topic modeling studies.

Author	**Country**	**The algorithm used**	**Data source**	**Purpose**	**Result**
Gozzi et al. [82]	Italy, USA, UK, and Canada	LDA	News articles, You-tube, Reddit, Wiki-pedia	Analyzing the public's internet response and media content during the pandemic	It has been noted that the public reaction was mainly oriented in line with the media coverage.

Liu et al. [87]	USA	LDA	Reddit	Comparing people's behavior changes, thoughts, confirmed cases, and deaths support from NLP	While positive changes were observed in mask thoughts, the words, "social distance", "hygiene", "mask", "cough", "cases", and "test" were frequently encountered.
Q. Liu et al. [88]	China	LDA	Wise-Search data-base	Examining media reports about the outbreak and health communica-tion patterns governed by the media	Nine main themes were identified, including popular prevention procedures, medical treatment, and global or local social/economic impacts.
Han et al. [89]	China	LDA Random forest algorithm	Weibo	Examining the texts on the relevant platform to find out what the public thinks	Seven main topics and 13 sub-themes were determined, topic extraction and classification model were put forward.

As a result of the scanned literature, it was observed that the researchers, by pulling the main themes of the epidemic in news, articles, or tweets, divided the perceptions into different clusters, and analyzed what perspectives were formed in the public. According to the research, LDA, which is one of the most effective algorithms in technically unsupervised machine learning applications, is frequently utilized as a data source to obtain public information. It has been discovered that there are studies that also benefit from different social media platforms and the articles are concentrated in the USA, and China, which is most affected by the epidemic. Consequently, it was emphasized that the public expects transparency from policymakers, they turn their focus to scientific information and are under the intense influence of social media. It has been commented that policymakers should decide on precautionary strategies, support communication regimes, and follow the process closely in this regard.

2.1.2.2. Sentiment analysis studies on the perception of COVID-19 in the public

Sentiment analysis is a computer tool of NLP that researchers use to identify the public's point of view on a topic [90]. With the use of sentiment analysis, people's emotions can be interpreted and the interpreted emotions can be used where necessary for any organization [91]. In this way, organizations can act according to the observed attitudes of people and take on a more agile structure. Sentiment analysis methods can determine positive, negative, neutral poles, anger, sadness, happiness emotions, or mental states regarding the target subject, theme, or interest aspects [5]. Techniques related to this aspect have played a key part in the studies carried out to inform policymakers and understand the emotional change-development of the public in the COVID-19 by attracting the attention of researchers.

Tsai and Wang [90] aimed to perform a sentiment analysis study with the help of Twitter data, and also, examined the relationship between mood changes in tweets about the pandemic and public health policies. In this way, they hoped to test the efficacy of analyzing public opinion on public health policies and events using Twitter data since the outbreak began. As a result of applying a data preprocessing strategy, they discovered a substantial link between mood changes in the collected data. Furthermore, by examining the data, they validated using Twitter data to monitor people's attitudes, as it is fast and requires little human effort.

Lwin et al. [92] aimed to analyze the worldwide emotional trends based on four main emotions, namely anger, fear, joy, and sadness, as well as to investigate the underlying causes of the emerging emotions in this direction. By examining tweets from the early days of the COVID-19, researchers have observed that fears related to shortages of medical supplies and testing for coronavirus have surfaced. Furthermore, they noted that the emotions of the public were shifting from the fear phase to the anger phase, the data collected in the emotion cluster about sadness was gathered around the anxiety of losing loved ones, also, the data collected in the emotion cluster about joy included feelings related to health and gratitude.

Pokharel [93] aimed to conduct a sentiment analysis to determine how people living in different countries coped with the pandemic situation based on 12 selected countries of the tweets sent between March 11, 2020, and March 31, 2020. Consequently, it was mentioned while it was observed that people generally exhibited a hopeful and positive perspective, their emotions such as disgust, fear, and sadness came to the fore. The Netherlands, USA, Switzerland, and France were found to have more anger and distrust than the other countries included in the investigation.

According to the tweets shared on Twitter, Dubey [94] aimed to analyze emotions and feelings, as well as to investigate the situation of cyber racism that increased during the epidemic. 16,000 tweets on the 11 April 2020 and 16 April 2020 date scales were selected as a sample, the public's feelings, and emotions regarding keywords such as "China Virus", "Wuhan Virus", and "Chinese Corona Virus" were observed. As a result, it was discovered that cyber racism increased, and the analyzed data generally contained negative emotions, especially negative emotions towards China, WHO, and ethnic Asians. Furthermore, it was discovered that profanity and rude comments were frequently utilized and that emotions such as grief, fear, rage, and contempt were prominent.

In order to ascertain the mental state of a people during their health period, Singh et al. [95], who received support from Twitter, performed a sentiment analysis with the help of the BERT model, using two different datasets, one from the world and one from India, by performing an application on tweets sent by users. With the tweets collected, it has been determined that people around the world are more negative about the epidemic than in India and they have a higher tendency to spread the negativity to other people.

Samuel et al. [79] tried to detect public awareness of the epidemic by using tweets. They provided insights into the evolution of fear overtime at the height of the epidemic in the USA. As a result, it was discovered that the pandemic caused the development of negative emotions such as fear in the public. Also, it is suggested that the methods presented could be used for the needed solutions and strategies to prevent the spread of feelings such as fear, panic, and hopelessness.

Researchers, focusing on understanding the psychological impact of the pandemic on people, thought that the best way for this was sentiment analysis, and focused on analyzing perceptions related to the epidemic by making use of different methods. By observing which emotions came to the fore and which emotions remained in the background, comments were made about the general course of the epidemic, as well as its future course. It has been determined in this direction that Twitter is frequently used to reach the real emotional changes of the whole society. It has been observed that emotions such as sadness, fear, anxiety, and insecurity are at the forefront in the current situation, also, it has been established that there are more hopeful and positive emotions for the future.

2.1.2.3. Topic modeling and sentiment analysis studies on the COVID-19

Aware of the advantageous aspects of topic modeling and sentiment analysis methods, the researchers thought that combining these two techniques could be more efficient in listening to the voice of the public and observing the process. In this direction, they carried out their applications by taking support from both topic modeling and sentiment analysis techniques in their studies. Thus, while revealing the main themes of the epidemic, they also had the opportunity to observe the developing moods of the public. As a result of the studies, they made suggestions to policymakers for the future and revealed the advantages of these two techniques with the help of social media. The studies are summarized in detail in **Table 2.6**.

Table 2.6: Studies supported by Twitter using topic modeling and sentiment analysis techniques.

Author	Keywords	Purpose	Result
Boon-Itt and Skunkan [96]	"Coronavirus", "covid 19", "2019-nCov", "covid-19", "symptoms", "epidemic", "pandemic"	Raising public awareness of the epidemic and revealing issues of concern shared by social media users	It has been argued that Twitter is a powerful social media tool and that COVID-19 can be divided into three stages based on symptoms/propensity to spread. People have also been seen to develop unfavorable attitudes regarding the disease in question.
Yin et al. [97]	"coronavirus", "covid", "covid 19", "Wuhan quarantine"	To present a new framework to analyze the emotional dynamics and topics about the epidemic from social media posts	It has been observed that positive thoughts are more dominant than negative thoughts in social media, also, it has been determined that topics such as "stay safe at home" are dominated by positive emotions, while topics such as "people's death" contain intensely negative emotions.

Al-Rawi et al. [98]	"women", "feminine", "gentleman", "masculine", "non-binary", "transgender", "bi-spirit", "gay", "lesbian"	To be able to determine the concerns and discourses of the public, considering gender in the process of the epidemic	While it has been shown that there are substantial discrepancies in the topics spoken between different genders, it has been emphasized that Twitter can be beneficial in establishing the communication line during epidemics.
Chen et al. [99]	"Chinese virus", "corona", "covid-19", "coronavirus"	Using Twitter to learn about the public's opinion of the pandemic	While negative thoughts were generally encountered, it was determined that tweets containing discussion mostly contained anger, while others were more realistic by reflecting analytical thoughts.
Hung et al. [100]	"Coronavirus", "stay at home", "epidemic", "social distancing", "scam", "be kind", "health", "heroes", "isolation", "school cancelled"	Analyzing and identifying sentiments related to the pandemic on Twitter	Five main themes were identified: emotional support, health environment, business economy, social change, and psychological support. Also, it was discovered that the states with the highest levels of negative emotions were New Mexico Wyoming, Florida, Alaska, and Pennsylvania, while the states with the most positive emotions were North Dakota, Vermont, Colorado, Utah, North Carolina, Tennessee.
Abd-Alrazaq et al. [101]	"corona", "2019-nCov", "COVID-19"	Identifying the main topics shared by Twitter users about the pandemic	4 main themes and 12 topics related to the main themes were determined is the origin of the virus; its impact on resources, people, countries, and the economy; its sources; approaches to limit infection risk, and it was determined that the opinions on 10 subjects were positive and the opinions on 2 subjects on the negative side.

Ebadi et al. [102] aimed to identify hidden issues by examining publication similarities and the change of emotions in the January 2020-May 2020 date range with the help of multiple data sources by using a structural topic modeling framework. As a result of this, it was discovered that the scope of available studies varies considerably and focuses on two main themes. It has been noted that the first has a wider range of epidemic-related issues, whilst the second concentrates on the smart systems required or used to diagnose COVID-19.

Oyebode et al. [103], by using six different social media platforms, namely Twitter, Facebook, YouTube, Archinect.com, PushSquare.com, and LiveScience.com, aimed to analyze the public's thoughts and comments regarding the COVID-19 with more than 1 million data. As a result of the research, the prominent keywords were determined, and the positive or negative emotion poles of these words were revealed. 34 negative themes including economic, educational, socio-political, and political keywords, also, 20 positive themes including other keywords were identified.

With the support of NLP techniques, Li et al. [104] aimed to examine the tweets of the COVID-19 period in terms of mental health by developing models that classify them according to anger, disgust, expectation, fear, and sadness, joy, confidence, and surprise. In this direction, they manually tagged 1,000 tweets in English, created them with an educational target, as a result, proposed an approach in order to be able to identify the issues that cause fear-sadness and analyze the emotional tendency.

After the curfews in Central Europe, Hanschmidt and Kersting [105] aimed to analyze the emotions of the public reflected on social media platforms and to examine the connection with emotions by revealing the main themes discussed. After the limits were enforced, there was a fall in the rate of happy emotions, as well as a decrease in despair and anxiety. Life in quarantine, infection-related difficulties, the influence of the pandemic on public and private life, and social contact restrictions were the four primary issues that came to the fore.

Raju et al. [106] aimed to put forward an NLP model with high classification accuracy based on the BERT model, which will support taking urgent measures to stop the proliferation of the epidemic by cleaning the misinformation on social media for health institutions. As a result, the proposed model can clean the comments found to contain misinformation from social media and report to policymakers about posts with false information. It is thought that in this way, the fake news circulating in the media about the epidemic can be prevented.

Lee [107] aimed to determine the topics around which people's thoughts were shared on Twitter about COVID-19, why these issues came to the fore, and how emotions developed during the epidemic. Between October 14, 2020, and November 10, 2020, 146,576 tweets taken from Twitter were analyzed, As a result, it became clear that despite the intense interest of people in COVID-19, they were unconcerned with the measures shared on social media to stop the epidemic in their daily lives.

Vydra and Kantorowicz [108] aimed to determine how useful social media data are for policymakers using latent semantic scaling and structural topic models. In this direction, in the relevant data, it has been determined which issues are discussed concerning possible policy by focusing on unemployment, early childhood care/education, and how these issues are shaped in the developing-changing economic, social, and policy change processes.

Focusing on different periods, M. S. Ahmed et al. [109] aimed to examine the sentiment of users and support trending topics using a dataset of tweets about the epidemic. As a result of the study, it was observed that emotions change after a certain time, different results that create diversity in social media communication were observed, and a model was presented to determine the sentiment dynamics for the most popular topics related to the pandemic.

In the early part of the pandemic, Mackay et al. [110] conducted an evaluation with the help of Facebook data, which included 438 posts and 26,774 comments. As a result, it has been seen that limited guiding principles may lead to the emergence of negative feelings in the public, also, it has been emphasized to the researchers that it is necessary to share the guiding information consistently to increase positive feelings in the public health.

Fazeli et al. [111] conducted a study when the epidemic first emerged, aiming to present a framework and present a multi-faceted analysis of the critical features reflected by the conversations on social media. With the support of data collected from Twitter and Reddit, a BERT-based framework has been made public. It has been commented that the developed framework can provide policymakers with insight that will not be costly but will provide scope throughout the pandemic, and in this context, the use of social media data will be efficient.

The authors, who wanted to reveal the public's perception of the pandemic, tried to put forward models that could guide by utilizing topic modeling and sentiment analysis techniques. With the models they put forward, they aimed to determine the different themes circulating among the people, as well as to reveal the mood of the people about the epidemic and make future suggestions. Thus, with the help of two NLP techniques, more comprehensive results were revealed, and both main themes and emotions were determined with sharp lines. At this point, the striking point is that social media platforms are predominantly preferred in studies, and it has been determined that Twitter is the most used social media platform to obtain clearer data. As a consequence, it has been observed that these two techniques are used separately in the articles, to the best of our knowledge, there hasn't been any research that combines these two methods and puts forward a model.

2.1.2.4. COVID-19 vaccine perception and topic modeling

Researchers aimed to reveal the attitudes of the public toward COVID-19 vaccines, which are thought to be a way to prevent COVID-19. In addition, they aimed to present studies that approached the event thematically and determined the views of the society to understand people's perspectives on both a possible vaccine and the current vaccine, in this way, they benefited from topic modeling techniques. While some researchers preferred to use Twitter as a data source, some researchers carried out their studies by pulling their data from different websites. These studies are summarized in **Table 2.7**.

Duraivel and Lavanya [112] attempted to look into the reasons for vaccine hesitancy, rejection, and resistance with the support of Reddit. Consequently, it was determined that the vaccine-hesitant gathered around the world, especially in the USA, also, the main themes were gathered around the fear of risks and side effects, lack of trust in policymakers, religious beliefs, mass observation theories, and the perception of vaccination as a priority to one-partyism.

Table 2.7: Studies supported by Twitter while using topic modeling techniques to detect vaccine perception.

Author	The algorithm used	Number of tweets	Purpose	Result
Jiang et al. [65]	LDA	100,209	Uncover themes found in the public's thinking about possible vaccines early in the COVID-19	It has been analyzed those public concerns have concentrated on broad themes concerning vaccines and that the views that emerged when a vaccine was not available were largely mixed with political debates.

S. Liu et al. [87]	LDA	2,678,372	Observing temporal changes by uncovering themes in tweets containing behavioral intentions, views, and attitudes towards the vaccine	A way for automating the process has been described to determine the public's views on vaccines, tweets are divided into categories such as attitudes and behavioral intentions, and each group is assigned 10 key themes. A large increase in tweets with positive behavioral goals was seen in December 2020.
Thurson [113]	LDA	31,925	Analyzing comments on social media about parents, children, vaccines, and COVID-19	Thoughts on parenting, vaccines, and COVID-19 were found on Twitter, also, 7 different categories were found: government, emotions, school, public health, Christmas, risk/safety, and parents/families.

In topic modeling studies, it has been observed that Twitter is generally used to pay attention to the voice of the public by using the LDA method, but besides Twitter, blogs, articles, and news sites are examined. The authors identified the main themes, made recommendations to policymakers, and examined the public's thoughts, confidence, and discourses against vaccination. It was discovered that vaccine rejection can be turned into acceptance with supported effects and information sharing. However, it has been concluded that far less than the number of vaccines required worldwide is still made to gain herd immunity, also, that vaccine rejection and hesitation are very high. In addition, while the researchers noticed parents' attitudes towards vaccination were hesitant, to determine the reasons behind this, topic modeling techniques were used with a focus on revealing the main themes, it was attempted to find out in this manner from which angles the parents looked at the situation and why their concerns arose.

2.1.2.5. Sentiment analysis studies on the COVID-19 vaccine

Focusing on analyzing and classifying text emotions, sentiment analysis can describe people's perspectives on public health policies for policymakers [90]. In this way, subjective opinions can be categorized within themselves and gain meaning. With this respect, sentiment analysis techniques attract the attention of the authors and have been widely used in examining the public acceptance of vaccines produced or being developed for the coronavirus. For example, Sandaka and Gaekwade [114] aimed to present a model to perform sentiment analysis on the tweets about the vaccine and to categorize the tweets as positive, negative, or neutral. As a result of the research, they discovered that the VADER model was the most successful in performing the sentiment analysis of the tweets posted by the public and gave the most appropriate results among the sentiment polarity scores. In addition, with time series analysis, they determined how daily mood changes fluctuate, with daily trend analysis, people's daily emotions, and with seasonal decomposition results, how the world reacts to the current situation. In addition, other studies about this topic in the literature are summarized in **Table 2.8**.

Table 2.8: Studies using sentiment analysis techniques to detect vaccine perception.

Author	**Time range**	**Data source**	**Purpose**	**Result**
Ahmed et al. [91]	16 February 2021- 5 April 2021	Surveys	Analyzing public opinion about the COVID-19 vaccine	It has been commented that strategic moves are needed to ensure that everyone understands and receives the COVID-19 vaccine.
Sattar and Arifuzzaman [115]	April 2020- May 2020	Twitter	Identifying vaccine insights and examining the public's perspective on safety measures after intake	It has been observed that people have a good attitude about vaccines and that this attitude persists after vaccination. By the end of July 2021, it is expected that 62.44% of the population will have received at least one dosage and 48% will be fully vaccinated.
Hussain et al. [5]	March 2020- November 2020	Face-book Twitter	Developing an artificial intelligence-based model by analyzing vaccine shares	The public is optimistic about vaccine development, trials, and efficacy, but concerned over safety, corporate control, and economic viability. Policymakers have been advised to employ artificial intelligence-assisted social media analysis to observe popular sentiment.
Gbashi et al. [116]	2 February 2020-5 May 2020	Twitter Google News	Performing systematic review for vaccine opinion	Contrary to popular belief, it was determined that the stance against the vaccine was positive and passive, also, it was emphasized that such an analysis could be successful in detecting potential attitudes and dominant poles.

In order to analyze the mood of the people against the vaccines to be made to prevent the epidemic, the authors benefited from different sentiment analysis techniques. It has been emphasized that states and policymakers should be more transparent and clearer to the public about the vaccine to prevent these concerns. In addition, most of the researchers have tried to put forward sentiment models that can analyze public opinion for possible vaccines, current vaccines, or they have tried to reveal the public perception clearly by trying to determine which is the most appropriate technique for data processing by using existing sentiment analysis techniques.

2.1.2.6. COVID-19 vaccine perception with topic modeling and sentiment analysis

The authors, who want to take the advantageous aspects of the topic modeling and sentiment analysis methods under the umbrella of NLP, received support from these techniques to determine the attitudes about the recently popular COVID-19 vaccine. In this way, they have introduced models to the literature that extract key themes and sentiments about vaccines. For example, by tackling different groups on Reddit, Wu et al. [117] sought to develop a clear understanding of the ideas about vaccines that raised key concerns. As a result of the study, it was determined that conspiracy theories predominated during the epidemic, each concern. As a consequence of the investigation, it was discovered that the sub-topic has its user group, users' anxiety differed according to time and topics, also, it was emphasized that communication efforts should be created according to specific needs. Other studies identified in the literature are summarized in **Table 2.9**.

Table 2.9: Studies using topic modeling and sentiment analysis techniques to analyze vaccine perception.

Author	Time range	Number of tweets	Purpose	Result
Lyu et al. [118]	11 March 2020-31 January 2021	1,499,421	To be able to identify the main themes and feelings discussed the vaccine	While positive sensitivity and dominant confidence about vaccines stand out, it has been commented that COVID-19 vaccines are more accepted than previous vaccines.

Kwok et al. [119]	January 2020-October 2020	31,100	To be able to deduct topics and feelings about the vaccine	While three main issues were identified, namely misunderstandings/ complaints regarding COVID-19 control, attitudes towards vaccination, and advocating infection control measures, it was concluded that two-thirds of the opinions were positive due to trust, and one-third were negative due to fear.
Shim et al. [1]	23 February 2021-22 March 2021	3,509	Analyzing, interpreting, and classifying tweets about the vaccine in Korea	It has been determined that vaccine hesitancy, which consists of fear, vaccine safety, influenza, period, and the degree of symptoms, is the most frequently discussed topic. It was discovered that as the number of cases increased, the shares tended to be negative.
Karami et al. [120]	1 November 2020-28 February 2021	185,953	Identifying the sentiment of tweets, pinpointing key issues, examining change, and revealing key points	Between November 2020 and February 2021, it was discovered that attitudes toward the vaccine shifted slightly to the favorable. Furthermore, it has been established that positive and negative posts have varied focuses, with topics ranging from immunizations to USA elections.

While the authors aimed to identify the main topics and feelings from the ideas, and discourses of the public using vaccine-related keywords, they preferred social media, especially Twitter. It has been observed that researchers mostly benefit from LDA, which is one of the unsupervised learning techniques of topic modeling. It was found that control measures against the vaccine were generally supported, and false information was rejected. Conspiracy theories have created a shield against the vaccination even though it has been noticed that positive emotions are not enough to reach the amount of vaccine needed to achieve herd immunity. It was emphasized to governments that they should pay attention to the people's voices, and also, be pioneers in supporting vaccine development and clinical management by creating a vaccine promotion plan.

CHAPTER 3

3. EXPERIMENTAL PART

The fundamental purpose of this research is to look into the perspectives of parents' attitudes and beliefs regarding getting their children vaccinated against coronavirus, which is critical for herd immunity, and to understand the reasons underlying these beliefs. In this regard, it is targeted to analyze the data set, which was taken from Twitter within certain limits, by applying topic modeling and sentiment analysis methods, with the support of the R software program.

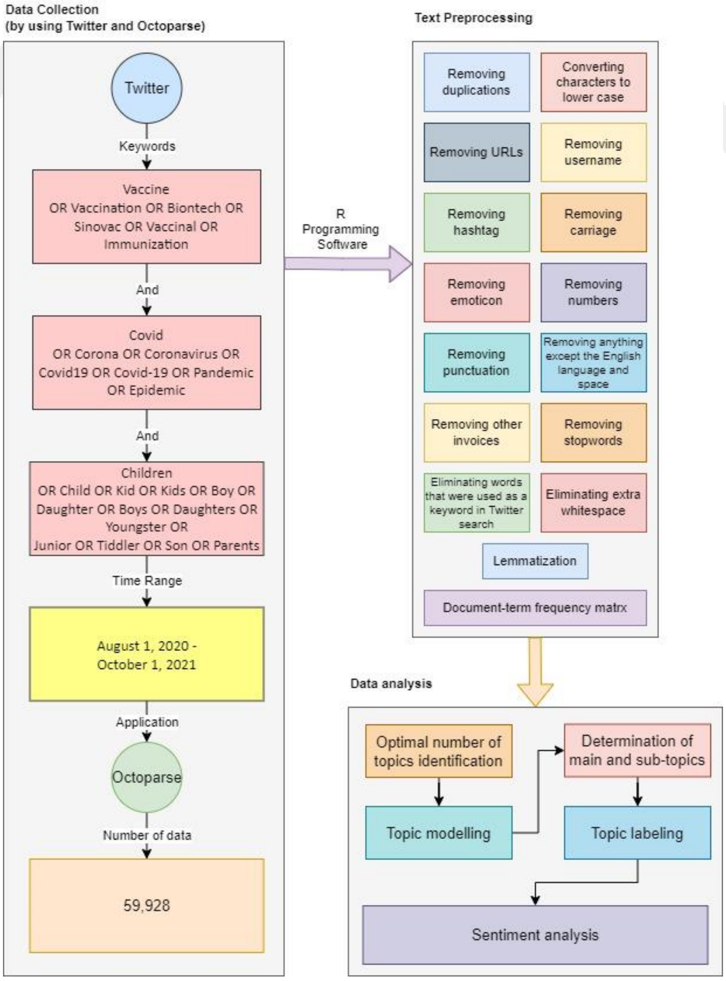

Figure 3.1: Summary of data collection process and analysis methods.

Accordingly, in this chapter, the data collection process and analysis methods are summarized (see **Figure 3.1**). The study consists of 3 steps in total. In step-1, data was gathered from Twitter with the help of the Octoparse web scraping tool by using selected keywords. In Step-2, the data was drawn into the R software program and the preprocessing process was carried out. In Step-3, data analysis was done. Topic clusters were extracted from the data set and labeled with the support of the chosen topic modeling method. Also, sentiment analysis was carried out to reveal the mood of the people belonging to the topic clusters.

3.1. Web Scraping and Dataset Generation

Many online platforms have emerged in recent years that connect various people, hence producing additional sources of information. Because there is so much information regarding different fields and topics created online, researchers collect input data for analysis from different social media platforms in their studies. At this point, web scraping comes into play, facilitating the process and helping to automatically extract the necessary information [121].

Web scraping is the process of gathering unstructured data from the internet and combining it with other data to create large-scale structured datasets. While there are numerous approaches to completing this task, it is usually done with the support of an application or software with the help of a browser extension. In this process, by writing an automatic program, data is requested by querying the web server, and the collected data is parsed to extract the necessary information [122]. With the aspects described, it provides a great advantage to researchers in order to easily collect unbiased information from the whole world and supports the collection of vast amounts of data with the necessary limitations.

Twitter is a social media platform where people freely share their views, bring together people from different regions, and have millions of information on almost every subject. For this reason, it is a platform preferred by researchers to create large data sets in the process of performing analysis on a specific topic. Based on the literature analysis, Twitter was chosen as the online venue for the construction of the data set in this study, and 59,928 tweets indicating families' perspectives on the coronavirus vaccine to be administered to their children were captured using the Octoparse software tool.

Octoparse is a piece of software that allows you to scrape any desired website offering a free plan up to a certain point to generate datasets in different formats. It is an easy-to-use yet powerful tool used to create the dataset [123]. Due to all the advantages, it provides, it was decided to use it to extract data from Twitter in this study, and it provided great benefits with the ease of use during the creation of the data set. In this direction, data was extracted from Twitter with the help of the Octoparse web scraping tool, using keywords determined in English in order to address the entire world, between August 1, 2020, and October 1, 2021, when the epidemic became a global problem and vaccine discussions became more intense.

Firstly, as a consequence of the literature review, the keywords to be searched on Twitter, which can pull tweets based on specified topics, were determined **(Table 3.1)**.

Table 3.1: Keywords used in an advanced search of Twitter.

Vaccine related keywords	COVID related keywords	Child-related keywords
• Vaccine • Vaccination • Biontech • Sinovac • Vaccinal • Immunization	• Covid • Corona • Coronavirus • Covid19 • Covid-19 • Pandemic • Epidemic	• Children • Child • Kid • Kids • Boy • Daughter • Boys • Daughters • Youngster • Junior • Tiddler • Son • Parents

Then, the following search lines were created by using the advanced search feature of Twitter with the determined keywords;

- (Vaccine OR Vaccination OR Biontech OR Sinovac OR Vaccinal OR Immunization) (Covid OR corona OR coronavirus OR Covid19 OR Covid-19 OR pandemic OR epidemic) (children OR child OR kid OR kids OR boy OR daughter

OR boys OR daughters OR youngster OR junior OR tiddler OR son OR parents) lang: en until:2021-10-01 since:2020-08-01 -filter: links -filter: replies

With the help of the written code, tweets that include the common intersection of a list of words linked to the 3 main topics that form the basis of the study, vaccine, covid, and child, also, tweets sent in English were determined within the specified date range. Furthermore, the search did not include tweets with links that were sent in response to another user. As a result of this procedure, a link was created and transferred to the Octoparse web scraping tool. By running the Octaparse, a flow was created within the program (see **Figure 3.2**), the columns containing the desired data were selected (see **Figure 3.3**) and the program was run.

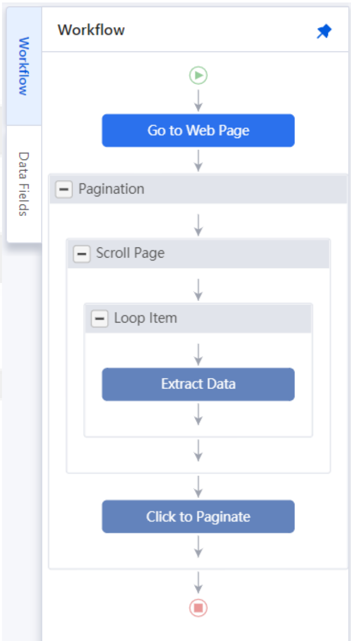

Figure 3.2: Created workflow on Octoparse.

#	Username	Tweet	Date
1	hoe-kage	I just wanna know why girl...	29 Eyl 2021
2	#MelnykOut	Relentless. Support. For. W...	29 Eyl 2021
3	Jacob Rossi	I'm one of the few people ...	29 Eyl 2021
4	Daniel Hatfield	Why parents should embr...	29 Eyl 2021

Figure 3.3: Columns containing the desired data.

By running the flow, a data set of 3 columns and 59,928 rows in total was created to be used in further analysis and exported in Excel form to do observation clearly on it. After this step, the created data set was transferred to the R program and a corpus was created for preprocessing and further analysis. The code which is available in Appendix A1 for pre-processing, topic modeling and sentiment analysis was utilized.

3.2. Pre-processing

R is a programming language and software environment for reporting, graphical representation, and statistical analysis [124]. It has been preferred for the analysis to be used in this study, as it provides different visual outputs, facilitates the interpretation of the outputs, and provides rapid processing of a large amount of data at the same time.

The data set created for the preprocessing process was transferred to R Studio and the preprocessing process was started. As a result, the information was cleaned and made ready for analysis. The cleaning procedure was carried out with the assistance of the functions, involving removing duplications, converting characters to lower case, removing URLs, removing usernames, removing hashtags, removing carriage, removing emoticons, removing numbers, removing punctuation, removing anything except the English language, and space, removing other invoices, removing stop words, eliminating terms that were utilized as a keyword in the Twitter search, eliminating extra whitespace, lemmatization and creating document-term frequency matrix.

The preprocessing of the input is of great importance in analyzes using NLP methods. This is the process of cleaning and preparing the input texts for analysis, and it is considered the initial stage in text classification. Data gathered from social media platforms, in particular, contains a lot of noise. There is a lot of data in datasets made using Twitter that does not contain meaningful information for analysis. By performing the preprocessing before the analysis, these data, which are called noise and have no benefit to the analysis, are removed from the data set [125].

In this context, necessary packages and libraries have been loaded into the R program (see **Figure 3.4**) to use the relevant functions for the preprocessing process. After this step, the data set was transferred to the program, then the column containing the tweets, that is, the information to be analyzed, was transferred to the corpus (see **Figure 3.5**). A corpus is a systematic collection of electronically recorded original texts that can be utilized to uncover language information that would otherwise go unnoticed [126]. Then, the preprocessing was carried out step by step.

```
"Installing related packages and libraries"
install.packages("tm")
library(tm)
library(textclean)
library(tokenizers)
install.packages("textstem")
library(textstem)
library(dplyr)
```

Figure 3.4: Installing related packages and libraries.

```
"Importing the dataset"
vaccination_data <- read.csv(file.choose(), header = T)
View(vaccination_data)

"Putting the data into a corpus for text processing"
tweet_corpus <- (VectorSource(vaccination_data$Tweet))
tweet_corpus <- Corpus(tweet_corpus)
```

Figure 3.5: Importing dataset and creating corpus.

3.2.1. Removing duplications

First, it is aimed to clean up the duplicate rows in the data set. As a result, the size of the data set will be reduced, and the data set will be cleansed of repeated information that will confound the analysis. The working logic of the code shown in **Figure 3.6** is based on detecting the repeating elements, separating the unique elements, and removing the repeating elements from the unique elements (see **Figure 3.7**). Also, with the assistance of the inspecting function, it has been observed that the code rows work correctly.

```
"Removing duplications"
removeDup <- function(x) unique(x)
tweet_corpus <- tm_map(tweet_corpus, removeDup)
inspect(tweet_corpus[1530:1540])
```

Figure 3.6: Removing duplications from the dataset.

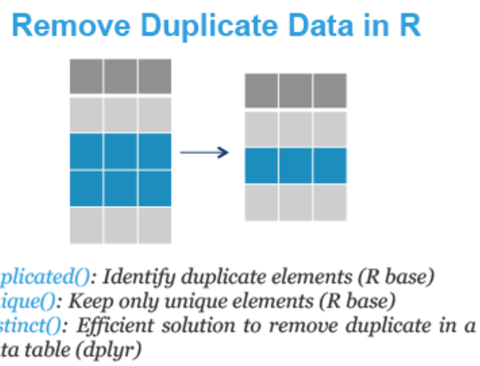

Figure 3.7: Working logic of the written code for removing duplications [127].

3.2.2. Converting characters to lower case

```
"Converting characters to lower case"
tweet_corpus <- tm_map(tweet_corpus, content_transformer(tolower))
inspect(tweet_corpus[1:131])
```

Figure 3.8: Converting characters to lower case in the dataset.

As a second step, all uppercase letters in the dataset were converted to lowercase using the code in Figure 3.8. The function for executing this step takes a letter as a function and returns the lowercase version and works based on this logic. R software is case sensitive, so converting it to lowercase helps with more robust results. **Table 3.2** shows the condition of the data set before and after the appropriate code has been executed, using sample rows for illustration.

Table 3.2: Before and after illustration of data while converting characters to lower case.

Before	After
WAKE UP #Nepal 12 YEAR OLD BOY AFTER THE COVID VACCINATION. You may not see the immediate reaction after taking the vaccine but spike protein in the vaccine is destroying your immune system. When people start to fall down, they will blame the variant not the vaccine.	wake up #nepal 12 year old boy after the covid vaccination. you may not see the immediate reaction after taking the vaccine but spike protein in the vaccine is destroying your immune system. when people start to fall down, they will blame the variant not the vaccine.
I now feel the need to quote tweet this because the \n@BoppyCompany\n newborn lounger was just recalled for EIGHT deaths out of >3 million units sold. Why canâ€™t we adults do the same for our children regarding vaccination? #foodforthought #COVID19 #VaccinesWork	i now feel the need to quote tweet this because the \n@boppycompany\n newborn lounger was just recalled for eight deaths out of >3 million units sold. why canâ€™t we adults do the same for our children regarding vaccination? #foodforthought #covid19 #vaccineswork

3.2.3. Removing URLs

URLs that did not make sense in the analysis were cleaned in the third phase. The links were not included in the search results for the Twitter advanced search, however, this step was taken to verify that the dataset was cleansed (see **Figure 3.9**).

```
"Removing URLs"
removeURL <- function(x){
  gsub("http[^[:space:]]*", "", x)
}
tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeURL))
inspect(tweet_corpus[1:131])
```

Figure 3.9: Removing URLs from the dataset.

3.2.4. Removing username

As a fourth step, usernames that did not contain meaningful information for the analysis were deleted from the tweets with the support of the written code (see **Figure 3.10**). **Table**

3.3 shows the condition of the data set before and after the appropriate code has been executed, using sample rows for illustration.

```
"Removing username"
removeUsername <- function(x){
  gsub("@[^[:space:]]*", "", x)
}
tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeUsername))
inspect(tweet_corpus[1:131])
```

Figure 3.10: Removing username from the dataset.

Table 3.3: Before and after illustration of data while removing username.

Before	After
wake up #nepal 12 year old boy after the covid vaccination. you may not see the immediate reaction after taking the vaccine but spike protein in the vaccine is destroying your immune system. when people start to fall down, they will blame the variant not the vaccine.	wake up #nepal 12 year old boy after the covid vaccination. you may not see the immediate reaction after taking the vaccine but spike protein in the vaccine is destroying your immune system. when people start to fall down, they will blame the variant not the vaccine.
i now feel the need to quote tweet this because the \n@boppycompany\n newborn lounger was just recalled for eight deaths out of >3 million units sold. why canâ€™t we adults do the same for our children regarding vaccination? #foodforthought #covid19 #vaccineswork	i now feel the need to quote tweet this because the \n\n newborn lounger was just recalled for eight deaths out of >3 million units sold. why canâ€™t we adults do the same for our children regarding vaccination? #foodforthought #covid19 #vaccineswork

3.2.5. Removing hashtag

In a subsequent phase, hashtags were cleared using the code shown in **Figure 3.11** to directly access people's ideas. **Table 3.4** shows the state of the tweets before and after the hashtags were removed.

```
"Removing hashtag"
removeHashtag <- function(x){
  gsub("#\\S+", "", x)
}
tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeHashtag))
inspect(tweet_corpus[1:131])
```

Figure 3.11: Removing hashtags from the dataset.

Table 3.4: Before and after illustration of data while removing hashtag.

Before	After
wake up #nepal 12 year old boy after the covid vaccination. you may not see the immediate reaction after taking the vaccine but spike protein in the vaccine is destroying your immune system. when people start to fall down, they will blame the variant not the vaccine.	wake up 12 year old boy after the covid vaccination. you may not see the immediate reaction after taking the vaccine but spike protein in the vaccine is destroying your immune system. when people start to fall down, they will blame the variant not the vaccine.
i now feel the need to quote tweet this because the \n\n newborn lounger was just recalled for eight deaths out of >3 million units sold. why canâ€™t we adults do the same for our children regarding vaccination? #foodforthought #covid19 #vaccineswork	i now feel the need to quote tweet this because the \n\n newborn lounger was just recalled for eight deaths out of >3 million units sold. why canâ€™t we adults do the same for our children regarding vaccination?

3.2.6. Removing carriage

As the sixth step, the carriages in the tweets and the carriages created by the software program during the transfer of the dataset to R studio were removed (see **Figure 3.12**). Thus, a cleaner appearance has been achieved.

```
"Removing carriage"
removeCarriage <- function(x){
  gsub("[\r\n]", "", x)
}
tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeCarriage))
inspect(tweet_corpus[1:131])
```

Figure 3.12: Removing carriage from the dataset.

3.2.7. Removing emoticon

Emoticons are graphical icons made up of punctuation marks, letters, and numbers that depict an emotion or sentiment [128]. Therefore, it does not carry any important information for further analysis. With the code written in **Figure 3.13**, it was cleared from the data set and tweets were made simpler.

```
"Removing emoticon"
removeEmoticon <- function(x){
  gsub("[^\x01-\x7F]", "", x)
}
tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeEmoticon))
inspect(tweet_corpus[1:131])
```

Figure 3.13: Removing emoticons from the dataset.

3.2.8. Removing numbers

```
"Removing numbers"
tweet_corpus <- tm_map(tweet_corpus, removeNumbers)
inspect(tweet_corpus[1:131])
```

Figure 3.14: Removing numbers from the dataset.

Because numbers don't contribute much information in the analysis, they're usually deleted at this point. The data is very dispersed in its raw state and therefore it becomes imperative to remove all other non-text impurities as in other preprocessing steps. In this context, the data in **Table 3.5** has been cleared with the support of the code in **Figure 3.14**.

Table 3.5: Before and after illustration of data while removing numbers.

Before	After
wake up 12 year old boy after the covid vaccination. you may not see the immediate reaction after taking the vaccine but spike protein in the vaccine is destroying your immune system. when people start to fall down, they will blame the variant not the vaccine.	wake up year old boy after the covid vaccination. you may not see the immediate reaction after taking the vaccine but spike protein in the vaccine is destroying your immune system. when people start to fall down, they will blame the variant not the vaccine.
i now feel the need to quote tweet this because the \n\n newborn lounger was just recalled for eight deaths out of >3 million units sold. why canâ€™t we adults do the same for our children regarding vaccination?	i now feel the need to quote tweet this because the newborn lounger was just recalled for eight deaths out of > million units sold. why cant we adults do the same for our children regarding vaccination?

3.2.9. Removing punctuation

```
"Removing punctuation"
tweet_corpus <- tm_map(tweet_corpus, removePunctuation)
inspect(tweet_corpus[1:131])
```

Figure 3.15: Removing punctuation from the dataset.

Punctuation marks make up a major portion of the dataset, and the analysis requires simply the text. They have been deleted because they have no significance in the analysis and as a result of this circumstance, the words may be fully revealed (**Figure 3.15** and **Table 3.6**).

Table 3.6: Before and after illustration of data while removing punctuation.

Before	After
wake up year old boy after the covid vaccination. you may not see the immediate reaction after taking the vaccine but spike protein in the vaccine is destroying your immune system. when people start to fall down, they will blame the variant not the vaccine.	wake up year old boy after the covid vaccination you may not see the immediate reaction after taking the vaccine but spike protein in the vaccine is destroying your immune system when people start to fall down they will blame the variant not the vaccine
i now feel the need to quote tweet this because the newborn lounger was just recalled for eight deaths out of > million units sold. why cant we adults do the same for our children regarding vaccination?	i now feel the need to quote tweet this because the newborn lounger was just recalled for eight deaths out of million units sold why cant we adults do the same for our children regarding vaccination

3.2.10. Removing anything except the English language and space

After removing the symbols, numbers, and punctuation marks from the data set, the code in the **Figure 3.16** was written, then, run to ensure that everything other than the English language and spaces are completely cleared.

```
"Removing anything except the English language and space"
removeNumPunct <- function(x) gsub("[^[:alpha:][:space:]]*", "", x)
tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeNumPunct))
inspect(tweet_corpus[1:131])
```

Figure 3.16: Removing anything except the English language and space from the dataset.

3.2.11. Removing other invoices

```
"Removing other invoices"
removeInvoice <- function(x){
  gsub("inv/[0-9]+/+[xvi]+/[xvi]+/[0-9]+", "", x, ignore.case = T)
}
tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeInvoice))
inspect(tweet_corpus[1:131])
```

Figure 3.17: Removing other invoices from the dataset.

A second function was created to supplement the previous step, and the dataset was freed from all unnecessary clutter (see **Figure 3.17**).

3.2.12. Removing stop words

```
"Removing stopwords"
tweet_corpus <- tm_map(tweet_corpus, removeWords, stopwords("english"))
inspect(tweet_corpus[1:131])
```

Figure 3.18: Removing stop words from the dataset.

Stop word is a term that refers to frequently used words that do not have any important meaning in the dataset. The contribution rate of these words, which are highly present in the data set, to the analysis is very low. These include words such as "a","the","if" etc [129]. Only the words that contribute sense to the tweets are crucial in the study, hence stop words must be removed during preprocessing. As a result, the code shown in **Figure 3.18** was produced, and the dataset's stop words were eliminated (**Table 3.7**).

Table 3.7: Before and after illustration of data while removing stop words.

Before	**After**
wake up year old boy after the covid vaccination you may not see the immediate reaction after taking the vaccine but spike protein in the vaccine is destroying your immune system when people start to fall down they will blame the variant not the vaccine	wake year old boy covid vaccination may see immediate reaction taking vaccine spike protein vaccine destroying immune system people start fall will blame variant vaccine
i now feel the need to quote tweet this because the newborn lounger was just recalled for eight deaths out of million units sold why cant we adults do the same for our children regarding vaccination	now feel need quote tweet newborn lounger just recalled eight deaths million units sold cant adults children regarding vaccination

3.2.13. Lemmatization

Lemmatization, in its broadest definition, is a process that groups inflected words into one basic form and thus simplifies it, closely related to stemming. With this process, words with the same root but different inflections can be looked at as a single item. The main goal is to reveal the basic dictionary form of the related words by eliminating the inflections and prefixes of the words in the data set [130], [131].

• Looking – (Lemmatization) -> Look

• Books – (Lemmatization) -> Book

• Talking – (Lemmatization) -> Talk

```
"Lemmatization"
tweet_corpus<-tm_map(tweet_corpus, textstem::lemmatize_strings)
inspect(tweet_corpus[1:131])
```

Figure 3.19: Lemmatization process for the dataset.

In this context, the data set was transformed into the form in **Table 3.8** with the aid of the code that has been written in **Figure 3.19**, in order to reveal the topic clusters more clearly and comprehensibly, especially during the topic modeling phase.

Table 3.8: Before and after illustration of data while lemmatization process.

Before	**After**
wake year old boy covid vaccination may see immediate reaction taking vaccine spike protein vaccine destroying immune system people start fall will blame variant vaccine	wake year old boy covid vaccination may see immediate reaction take vaccine spike protein vaccine destroy immune system people start fall will blame variant vaccine
now feel need quote tweet newborn lounger just recalled eight deaths million units sold cant adults children regarding vaccination	now feel need quote tweet newborn lounger just recall eight death million unit sell cant adult child regard vaccination

3.2.14. Eliminating words that were used as a keyword in Twitter search

The words used in an advanced search on Twitter have been eliminated from the data arranged in such a way that they don't interfere with clustering in the topic modeling section. Also, by looking at the output several times, some words that don’t make sense to the model have been removed. Thus, the density and clustering caused by the use of tweets consisting of these words have been prevented (see **Figure 3.20** and **Table 3.9**).

```
"Eliminating words that were used as a keyword in Twitter search"
tweet_corpus <- tm_map(tweet_corpus, removeWords,c( "vaccine", "vaccination", "biontech","sinovac", "vaccinal", "immunization",
                                                    "covid", "corona", "coronavirus", "pandemic", "epidemic", "children",
                                                    "child", "kid", "boy", "daughter", "youngster", "junior", "tiddler", "son", "parent",
                                                    "vaccinate","get","just","today","now","good","today","like","can","day","people","see",
                                                    "take", "still","give","make","know","tell","say","year","one"))
inspect(tweet_corpus[1:131])
```

Figure 3.20: Eliminating words that were used as a keyword in a Twitter search from the dataset.

Table 3.9: Before and after illustration of data while eliminating words that were used as a keyword in Twitter search.

Before	**After**
wake year old boy covid vaccination may see immediate reaction take vaccine spike protein vaccine destroy immune system people start fall will blame variant vaccine	wake old may immediate reaction spike protein destroy immune system start fall will blame variant
now feel need quote tweet newborn lounger just recall eight death million unit sell cant adult child regard vaccination	feel need quote tweet newborn lounger recall eight death million unit sell cant adult regard

3.2.15. Eliminating extra whitespace

```
"Eliminating extra whitespace"
tweet_corpus <- tm_map(tweet_corpus, stripWhitespace)
inspect(tweet_corpus[1:131])
```

Figure 3.21: Eliminating extra whitespaces from the dataset.

With the support of the function that helps to remove extra spaces from a text document, the extra spaces that are already present in tweets and took place as a result of pre-processing steps are eliminated. Thus, multiple space characters are collapsed into a single space. Also, by removing the extra spaces in the data, the document matrix can be easily created. Because it's possible that the extra space, or any sign in the space, will be taken into account in the document term matrix (see **Figure 3.21** and **Table 3.10**).

Table 3.10: Before and after illustration of data while eliminating extra whitespaces from the dataset.

Before	**After**
wake old may immediate reaction spike protein destroy immune system start fall will blame variant	wake old may immediate reaction spike protein destroy immune system start fall will blame variant
feel need quote tweet newborn lounger recall eight death million unit sell cant adult regard	feel need quote tweet newborn lounger recall eight death million unit sell cant adult regard

After this stage is completed, it's time to export the processed data to CSV for analysis and generate a document term matrix for analysis. For future usage, the preprocessed data is exported to a CSV file (see **Figure 3.22**).

```
"Extracting preprocessed data into csv file for further usage"
dataframe <- data.frame(text=sapply(tweet_corpus, identity),
                        stringsAsFactors=F)
df <- data.frame(text = get("content", tweet_corpus))
head(df)
write.csv(df, 'my.csv')
```

Figure 3.22: Extracting preprocessed data into a CSV file for further usage.

3.2.16. Creating document-term matrix

The frequency of terms in a group of documents is represented by a document-term matrix, which is a mathematical matrix. A document-term matrix's rows correspond to the documents in the collection, while the columns relate to the terms [132], [133]. Natural language processing and computational text analysis both benefit from it. As the final step of the preprocessing to build the basis for topic modeling, the document term matrix was built with the help of the code in Figure 3.23.

```
"Creating document-term matrix"
dtm <- DocumentTermMatrix(tweet_corpus)
```

Figure 3.23: Creating document-term matrix.

3.3. Topic Modelling

As the world has become more digital, the number of people and platforms on social media using these platforms is increasing rapidly. With this rapid increase, the information shared on social media provides added value and relevance both to policymakers in the decision-making process and to researchers on the academic side in the process they carry out their studies, so the analysis of information flows comes to the fore. Combining text analysis tools with social media is seen as a good strategy to make better decisions since it gives a communication-oriented decision support system [134]. However, accessing the huge amount of data that such platforms provide and extracting meaningful parts from the large amount of information obtained is a difficult task. Therefore, automatic, new methods like topic modeling have emerged to help analyze, understand, and organize this data in social media.

Topic modeling is an unsupervised machine learning method frequently encountered in sciences such as NLP, information retrieval, and text mining, which creates a statistical model and is used to find the number of words frequently used in textual information [135]. The methods under this heading assume that meanings are composed of word sets that describe related topics of conversation while at the same time accepting meanings as relational [134]. Using this technique, decision-makers observe frequently used word patterns in data sets to obtain meaning, thus revealing the data's hidden information, that is, the topic of the data, and labeling them accordingly (see **Figure 3.24**). Through analysis, researchers can discover hidden themes in the collection, group documents into discovered themes, and summarize documents by topic [102]. Unsupervised topic models can summarize, understand, analyze, and organize information in large data sets.

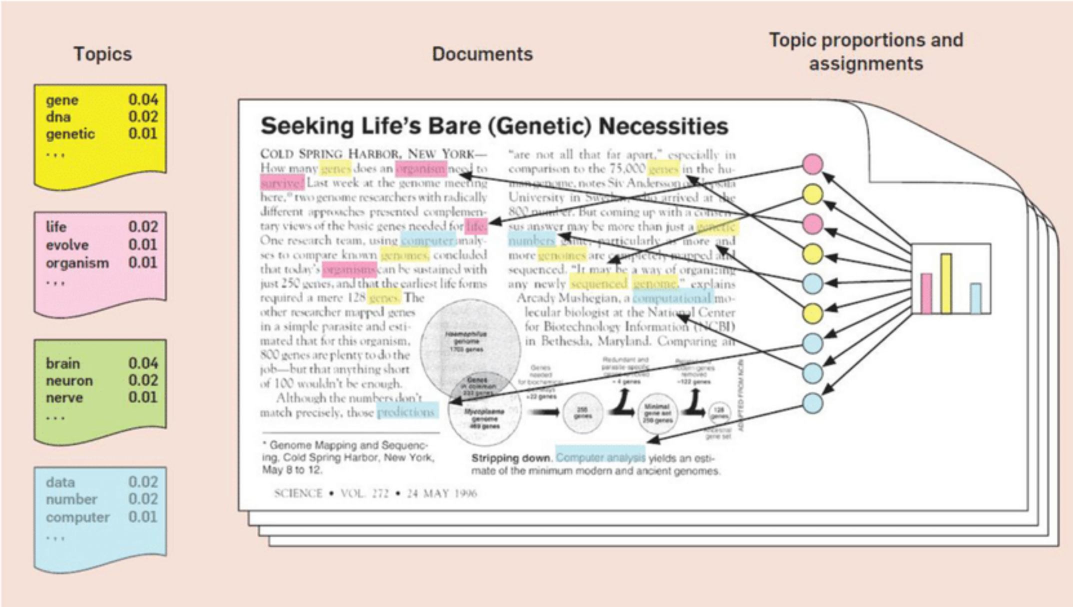

Figure 3.24: Logic behind the topic modeling [136].

Matrix decomposition approaches, which look for a low-dimensional representation of the data by factoring into low-order matrices, and probabilistic topic modeling methods, which look for generative statistical models, are the two types of topic modeling methods. The most popular analysis methods can be summarized as follows;

- LSA using SVD
- PLSA, which replaces LSA based on a polynomial model with a probabilistic approach
- NMF as a topic model that preserves the non-negative nature of the data
- LDA, the most popular thread model in use with NMF at the moment

LSA and NMF are parsing-based methods. PLSA straddles between being a parsing-based method and being a generative one. LDA is a generative model [137]. In this study, topic modeling was used to reveal topic clusters and analyze the most talked words about topics due to the diversity of information found on Twitter and the size of the dataset created. Support was also obtained from the LDA algorithm, which is claimed to be the simplest form of topic modeling and is observed to be frequently preferred in the literature.

3.3.1. Latent Dirichlet Allocation (LDA)

LDA is a generative statistical model that explains why some parts of a dataset are similar to each other but is the most popular topic modeling and analysis term, which assumes that there are certain topics known as fixed word distributions for the entire document collection [138]. It works by assuming that the topics make up the document and that each topic creates the words that make up the document, based on the probability distribution [139], [140]. LDA is widely regarded as a useful method for uncovering and exploiting a huge dataset's hidden thematic structure [141]. The steps used in LDA to identify the topic can be summarized as shown in **Figure 3.25** [142].

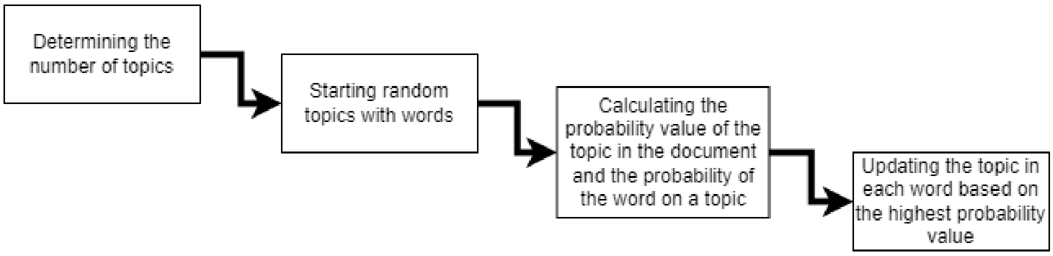

Figure 3.25: Steps of LDA.

The distribution used to determine the distribution of topics per document is known as the Dirichlet distribution. Then, in the generative process for the LDA method, the results from this distribution are used to allocate words in the document to different topics. While documents are observable in LDA, topics, distribution of topics per document, and classification of each word by topics per document are hidden constructs. This is where the algorithm got its name [143].

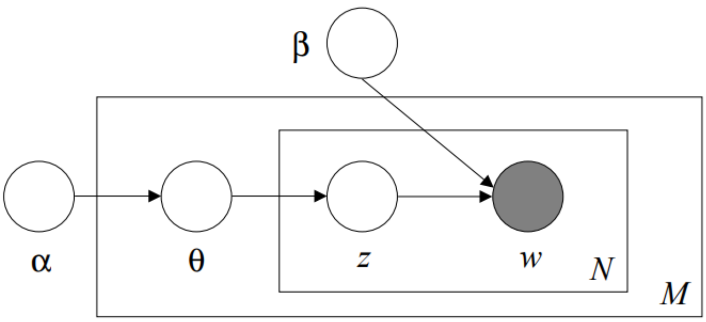

Figure 3.26: Graphical model representation of LDA [143].

In **Figure 3.26**, the graphical model representation of LDA was given. Replicas are represented by the boxes, which are "plates”. The outer plate represents documents, while the inner plate reflects the document's repeated selection of topics and words. In **Figure 3.26** [144];

α gives the distribution of topics per document, β gives the word distribution per topic,

Θ indicates the distribution of topics for a particular document,

z is the assigned topics for each word,

w are observed words.

The α and β parameters are sampled once during system creation. The Θ parameter is sampled for each document in the system. Also, the logic of the method can be supported by **Figure 3.27** for better presentation.

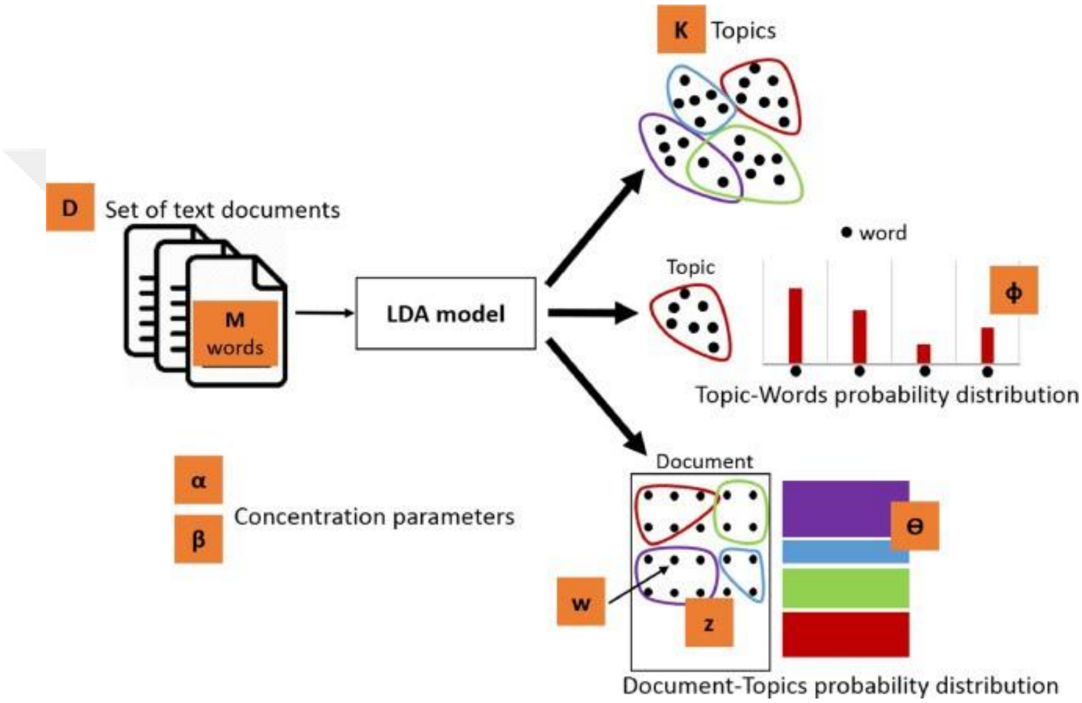

Figure 3.27: Logic of the LDA method [145].

The LDA algorithm, as mentioned before, is based on a statistical basis. However, a normal statistical model assumes that for each of the vocabulary words, there is a probability of that word occurring. LDA adds a layer of complexity to this system. LDA counts a list of topics (k). Each document (m) is a probability distribution over k topics. Each topic is a probability distribution over all the terms in our vocabulary (V). It accepts that every word has various possibilities of appearing in every topic. The exact probability formula that LDA uses and creates a document is explained in the following [146];

- $P(W, Z, \Theta, \varphi; \alpha, \beta)$
  - Total probability of the LDA model

- $\prod_{i=1}^{K} P(\varphi_i, \beta)$
  - Dirichlet distribution of topics over terms: for each topic i amongst K topics, what is the probability distribution of words for i.
- $\prod_{j=1}^{M} P(\Theta_j; \alpha)$
  - Dirichlet distribution of documents over topics: for each document j in the corpus of size M, what is the probability distribution of topics for j.
- $\prod_{t=1}^{N} P(Z_{j,t} \mid \Theta_j) P(W_{j,t} \mid \varphi_{Z_{j,t}})$
  - Probability of a topic appearing given a document and the likelihood of a word appearing given a topic: how likely is it that certain topics, Z, appear in this document, and then how likely is that certain words, W, appear given those topics.

Based on this informations, the following formula was appearing in the use of the LDA method;

$$P(W, Z, \Theta, \varphi; \alpha, \beta) = \prod_{i=1}^{K} P(\varphi_i, \beta) \prod_{j=1}^{M} P(\Theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t} \mid \Theta_j) P(W_{j,t} \mid \varphi_{Z_{j,t}}) \qquad \textbf{(1)}$$

The first two sums in the formula contain symmetrical Dirichlet distributions, which are called preliminary probability distributions for topics and documents. The third total includes two polynomial distributions, one on words and one on topics. By using the final probability formula, the same word distribution obtained in the original documents is tried to be created [146].

In this study, to cluster the topics, the LDA algorithm, which is simple to use and one of the most popular topic modeling methods in recent years, was chosen and topic clusters were developed with the help of the code written in R. In order to be able to observe and evaluate more easily, additional visualization functions were supported.

```
"Collapsing matrix by summing over columns"
dtm <- removeSparseTerms(dtm, sparse = 0.95)
frequency <- colSums(as.matrix(dtm))
"Length should be total number of terms"
length(frequency)
"Creating sort order (descending)"
ord <- order(frequency, decreasing = TRUE)
"Listing all terms in decreasing order of freq and write to disk"
frequency[ord]
write.csv(frequency[ord], "word_freq.csv")
"Removing the 0 rows"
raw.sum=apply(dtm,1,FUN=sum)
dtm=dtm[raw.sum!=0,]
```

Figure 3.28: Finding frequencies of words in descending order.

In this context, firstly, sparse terms in the matrix were cleaned to reduce the vectorial size of the data set, as shown in **Figure 3.28**. Then the matrix was narrowed by summing the columns, and descending order of frequencies was created. Afterward, the frequencies of the terms were recorded and all rows with a value of 0 were removed from the matrix. Thus, the model was made ready to create topic clusters.

```
"Creating model with 4 topics"
k=4
seed=1234
lda_fit <- LDA(dtm, k=k, control=list(seed=seed))
lda_fit@alpha
topics(lda_fit, k)
terms(lda_fit, 5)

"Tagging the docs to topics and topics to terms"
lda_fit.topics <- as.matrix(topics(lda_fit))
write.csv(lda_fit.topics,file=paste("docstotopics", k, "DocsToTopics.csv"))

lda_fit.terms <- as.matrix(terms(lda_fit,4))
write.csv(lda_fit.terms, file=paste("topicstoterms", k, "DocsToTerms.csv"))
lda_fit.terms[1:5,]
```

Figure 3.29: Creating topic clusters.

In order to evaluate the data set, it was decided to reflect and examine 4 topic clusters and the 5 words that make up these 4 topic clusters in the results. In this context, the code-written topic clusters in **Figure 3.29** were created. The results were recorded as CSV files, which were later used in sentiment analysis.

```
"Converting corpus for further manipulation"
tdm <- TermDocumentMatrix(tweet_corpus)
tdm <- removeSparseTerms(tdm, sparse = 0.95)
m <- as.matrix(tdm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)

"Plotting the most frequent words"
barplot(d[1:10,]$freq, las = 2, names.arg = d[1:10,]$word,
        col ="lightblue", main ="Most frequent words",
        ylab = "Word frequencies")

"Generating word cloud"
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 5,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
findFreqTerms(tdm, lowfreq = 4)

"Clustering associated words"
tdm2 <- removeSparseTerms(tdm, sparse = 0.95)
m2 <- as.matrix(tdm2)
distMatrix <- dist(scale(m2))
fit <- hclust(distMatrix, method = "ward.D2")
plot(fit)
rect.hclust(fit, k=4)
```

Figure 3.30: Visualization and analysis stage.

The corpus obtained for subsequent uses was transformed. The results were examined with different visualizations, and the process of cluster formation and labeling of the data was started (see **Figure 3.30**).

```
"Plotting a frequency graph based on topics' number"
top_terms_by_topic_LDA <- function(input_text,
                                   plot = T,
                                   number_of_topics = 4)
{

  topics <- tidy(lda_fit, matrix = "beta")

  top_terms <- topics  %>%
    group_by(topic) %>%
    top_n(5, beta) %>%
    ungroup() %>%
    arrange(topic, -beta)

  if(plot == T){
    # plot the top 5 terms for each topic in order
    top_terms %>% # take the top terms
      mutate(term = reorder(term, beta)) %>% # sort terms by beta value
      ggplot(aes(term, beta, fill = factor(topic))) + # plot beta by theme
      geom_col(show.legend = FALSE) + # as a bar plot
      facet_wrap(~ topic, scales = "free") + # which each topic in a seperate plot
      labs(x = NULL, y = "Beta") + # no x label, change y label
      coord_flip() # turn bars sideways
  }else{
    # if the user does not request a plot
    # return a list of sorted terms instead
    return(top_terms)
  }
}

top_terms_by_topic_LDA(tweet_corpus, number_of_topics = 4)
```

Figure 3.31: Plotting a frequency graph.

In order to observe the found topic clusters more easily, the code written in **Figure 3.31** was supported, so bar plots were drawn according to the probability of the words belonging to the topics. At this point, after the topic clusters were found with the support of LDA, the sentiment analysis part was started to determine attitudes and ideas more clearly.

3.4. Sentiment Analysis

Sentiment analysis techniques aim to reveal the emotions that exist for any topic in a text. Thus, it generally categorizes texts as positive, negative, or neutral based on emotion scores [147]. In short, it's the process of quickly reporting and making sense of massive amounts of data using software systems, with burgeoning social networks at the forefront. With the feature of sentiment analysis to question whether the texts have positive, negative, or neutral content in general, the attitude of the individuals concerned about the relevant topic is determined [148]. Opinions and feelings significantly affect the decision-making process in every field.

Twitter sentiment analysis is a prominent concept within the framework of computational linguistics sentiment analysis. Approaches to sentiment analysis identify and evaluate the views expressed in the text with the help of automated methods [149]. The emotion can be found in the tweet or shared by individuals to provide useful indicators for different purposes. Sentiment analysis is a natural language processing approach used to measure an idea or emotion expressed in a series of tweets, which was applied in this study (see **Figure 3.32**) [150].

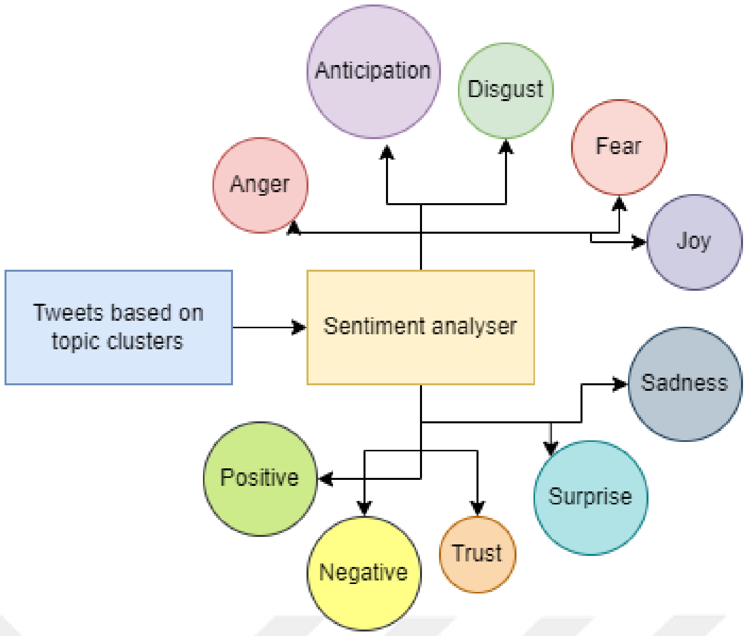

Figure 3.32: Visualization of the sentiment analysis process.

The emotional moods of the topic clusters were determined using sentiment analysis methods in this study. To accomplish the sentiment analysis, a code draft was written and executed for four different topic clusters. As the first step, the packages and libraries that will support the operation of the code have been installed as shown in **Figure 3.33** to be able to conduct sentiment analysis and use the necessary visualizations for the analysis.

As the next step, sentiment analysis was conducted using the draft code created for each topic cluster. The code run for topic 1 will be shared as an example. After the packages and libraries were loaded, the CSV file was used, in which the labels performed in the topic modeling were transferred. Different CSV files consisting of tweets hosted by clusters were created for 4 topics to obtain sentiments.

```
"Installing packages"
install.packages("tm")  # for text mining
install.packages("SnowballC") # for text stemming
install.packages("wordcloud") # word-cloud generator
install.packages("RColorBrewer") # color palettes
install.packages("syuzhet") # for sentiment analysis
install.packages("ggplot2") # for plotting graphs
install.packages("RCurl", repos = "http://cran.us.r-project.org")
install.packages("httr", repos = "http://cran.us.r-project.org")
install.packages("syuzhet", repos = "http://cran.us.r-project.org")
install.packages("plotly")

"Loading libraries"
library(NLP)
library("tm")
library("SnowballC")
library("wordcloud")
library(wordcloud2)
library("RColorBrewer")
library("syuzhet")
library(stringr)
library("ggplot2")
library(reshape2)
library(janeaustenr)
library(dplyr)
library(stringr)
library(tidytext)
library(sentimentr)
library(plyr)
library(RCurl)
library(httr)
library(topicmodels)
library(tidyr)
```

Figure 3.33: Installing related packages and loading libraries.

With the support of the code used in the topic modeling and shared in Figure 3.32, the data from the CSV created based on the topics were transferred to RStudio.

```
"Importing the dataset"
vaccination_data_t1 <- read.csv(file.choose(), header = T)
View(vaccination_data_t1)
```

Figure 3.34: Importing the dataset for topic 1.

After the data was transferred to the program, the sentiment scores of the tweets were calculated. Thus, the basis of the analysis was formed and observed, then, other stages were passed. As seen in **Figure 3.35**, these code rows are written to observe the process in the CSV's first 10 lines. Anger, anticipation, disgust, fear, joy, sadness, surprise, trust, and negative and positive emotional states were scored with the support of the words contained in the tweets. However, it is very difficult to interpret from here, so the calculated scores were visualized and interpreted in the following stages.

```
"Counting emotions from dataset with the help of relatd library"
d <- get_nrc_sentiment(vaccination_data_t1$Tweet)
head (d,10)
```

Figure 3.35: Counting emotions from the dataset with the help of a related library.

Code lines shared in **Figure 3.36** were created to observe and interpret the general score status of all tweets with the help of the calculated scores. Thus, a plot was drawn by considering the emotion scores of all the data in the data set.

```
"Transposing"
td<-data.frame(t(d))
"Computing column sums across rows for each level of a grouping variable"
td_new <- data.frame(rowSums(td[1:5802]))
"Transformation and cleaning"
names(td_new)[1] <- "count"
td_new <- cbind("sentiment" = rownames(td_new), td_new)
rownames(td_new) <- NULL
td_new2<-td_new[1:8,]
"Plotting one - count of words associated with each sentiment"
quickplot(sentiment, data=td_new2, weight=count, geom="bar", fill=sentiment,
          ylab="count")+ggtitle("Sentiment scores of tweets")
```

Figure 3.36: Plotting one count of words associated with each sentiment.

In addition, another plot was drawn to see the percentages and sequential form of the emotion scores in the general situation, also the interpretations were supported by this plot (see **Figure 3.37**).

```
"Plotting two - count of words associated with each sentiment, expressed as a percentage"
barplot(
  sort(colSums(prop.table(d[, 1:8]))),
  horiz = TRUE,
  cex.names = 0.7,
  las = 1,
  main = "Emotions in Text", xlab="Percentage"
)
```

Figure 3.37: Plotting two - counts of words associated with each sentiment, expressed as a percentage.

After examining the emotional states of the tweets in detail, the negative score was subtracted from the positive score to understand whether the tweets were closer to the positive or negative side. In addition, to interpret the scores on a topic basis, the output obtained here is saved in a CSV file (see **Figure 3.38**). The code in **Figure 3.39** was used to make sense of the incoming output, and also, to see the net scores of the emotional states based on emotion and general score. At the same time, the rows of code in **Figure 3.40** and **Figure 3.41** were written to enrich the visualization and to see the place occupied by the emotion scores in the total of the data set.

```
d$score<-d$positive-d$negative
d[1:10,]

write.csv(x=d,file="C:/Users/elifd/Desktop/2021-2022/YL/Tez/Kod ve deneme belgeler/skor.csv")
```

Figure 3.38: Calculating positivity and negativity scores of tweets.

```
"Calculating tweet scores in terms of different emotions"
tweetscore <- colSums(d[,])
print(tweetscore)
```

Figure 3.39: Calculating tweet scores in terms of different emotions.

```
"Emotion classification & positive and negative sentiments"

nrc_average <- apply(d,2,mean)
nrc_average
sentisum <- colSums(d)
sentisum
Lb <- paste(names(sentisum), ",", sentisum)
pie(sentisum[1:10],col=brewer.pal(8,'Dark2'), labels=Lb,
    main="Emotions and Sentiment nrc Scores", cex=0.8, cex.main=2)
```

Figure 3.40: A different illustration for emotion classification.

```
emotions <- get_nrc_sentiment(vaccination_data_t1$Tweet)
emo_bar = colSums(emotions)
emo_sum = data.frame(count=emo_bar, emotion=names(emo_bar))
emo_sum$emotion = factor(emo_sum$emotion, levels=emo_sum$emotion[order(emo_sum$count, decreasing = TRUE)])
# Visualize the emotions from NRC sentiments
library(plotly)
p <- plot_ly(emo_sum, x=~emotion, y=~count, type="bar", color=~emotion) %>%
  layout(xaxis=list(title=""), showlegend=FALSE,
         title="Emotion Type for vaccination")
p
```

Figure 3.41: Plotting the emotion type for visualization.

As a final step, a word cloud was established, and it was decided which emotions mostly contain which words. Thus, it was ensured that the feelings, attitudes, and ideas assigned to the topic were tied to a definite conclusion. To be able to implement this step, which will conclude the analysis, the code in **Figure 3.42** was run.

```
wordcloud_tweet = c(
  paste(vaccination_data_t1$Tweet[emotions$anger > 0], collapse=" "),
  paste(vaccination_data_t1$Tweet[emotions$anticipation > 0], collapse=" "),
  paste(vaccination_data_t1$Tweet[emotions$disgust > 0], collapse=" "),
  paste(vaccination_data_t1$Tweet[emotions$fear > 0], collapse=" "),
  paste(vaccination_data_t1$Tweet[emotions$joy > 0], collapse=" "),
  paste(vaccination_data_t1$Tweet[emotions$sadness > 0], collapse=" "),
  paste(vaccination_data_t1$Tweet[emotions$surprise > 0], collapse=" "),
  paste(vaccination_data_t1$Tweet[emotions$trust > 0], collapse=" ")
)

corpus = Corpus(VectorSource(wordcloud_tweet))
tdm = TermDocumentMatrix(corpus)

# convert as matrix
tdm = as.matrix(tdm)
td_new <- cbind("sentiment" = rownames(tdm), tdm)
colnames(tdm) = c('anger', 'anticipation', 'disgust', 'fear', 'joy', 'sadness', 'surprise', 'trust')
comparison.cloud(tdm, random.order=FALSE,
                 colors = c("#00B2FF", "red", "#FF0099", "#6600CC", "green", "orange", "blue", "brown"),
                 title.size=1, max.words=250, scale=c(2, 0.4),rot.per=0.4)
```

Figure 3.42: Plotting word cloud based on the titles of emotions.

With the preprocessing, topic modeling, and sentiment analysis, the time came to interpret the output of the sequence analysis and to finalize the study by looking at the model's output.

CHAPTER 4

4. RESULTS AND DISCUSSION

In this thesis, a model that provides the general attitude of families against the coronavirus vaccine in children was presented by analyzing tweets that contain the ideas of parents against the vaccine for their children. While establishing the model, analysis methods were used under the umbrella of NLP. Therefore, the conclusion part can be considered as two stages (see **Figure 4.1**). The first is the establishment of topic clusters by applying the LDA method, one of the topic modeling techniques that reveal the most talked about topics. The second one is the stage of performing sentiment analysis, which reveals attitudes by labeling the generated topic clusters to the data.

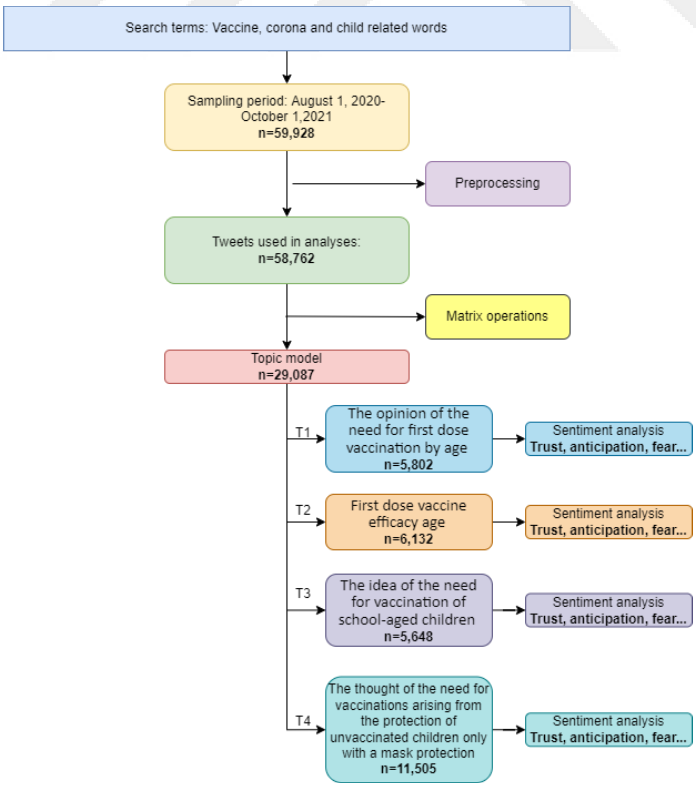

Figure 4.1: General Flow of Result.

4.1. Topic Modelling Results

LDA method, which is one of the topic modeling methods, was used to reveal the topic clusters and to observe the most frequently mentioned words. In this context, with the help of the web scraping tool, 59,928 tweets were retrieved as a result of an advanced search on Twitter. The captured data were transferred to the RStudio software program and analyzed by developing the LDA-based code. First, it was discovered that when the terms were organized according to their frequency in the data set (see **Figures 4.2** and **4.3**), the clearest question mark that arose in the thoughts of the families regarding the vaccine was related to their children's school status. Apart from this, the discussions about the first vaccination dose given to children were prominent in the tweets posted across the world in the English language. It was determined that the other words came from the standpoint of a desire, need, and thought for the vaccine. Also, the majority of interpretations are about the age of the children, the effectiveness of the vaccine, and its relationship with the mask.

```
> frequency[ord]
school  first   dose   want    old   need    age   mask   work  think
  6978   6511   4410   4201   4044   3972   3742   3379   3274   3266
```

Figure 4.2: Frequency of words in the dataset.

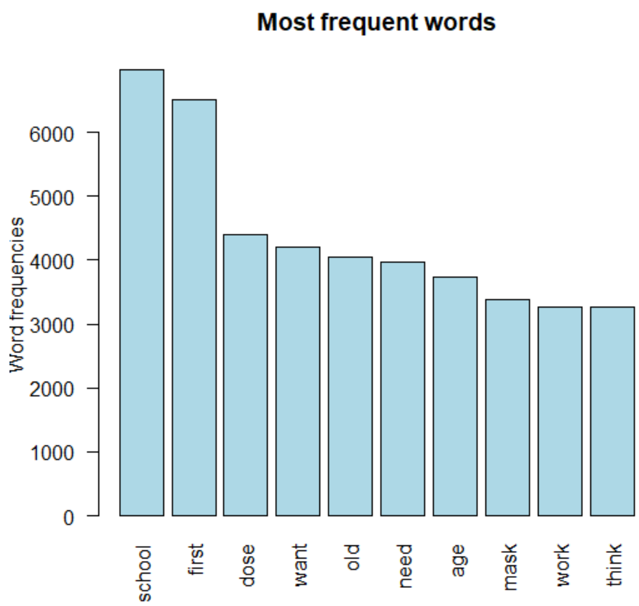

Figure 4.3: Frequency of words in the dataset on bar plot.

In addition, by using different visualizations such as word cloud beside the bar plot, prominent words are highlighted, and interpretations are facilitated (see **Figure 4.4**).

Figure 4.4: Word cloud with most frequent words.

A clustering diagram was created to see which words should be grouped to consolidate the topic clusters that are going to be determined, and the result is shared in **Figure 4.5**. This clustering diagram, called a dendrogram, uses the distance between 2 probability vectors to examine whether the terms are closely related. When the dendrogram created is examined, it is clear that the most frequent shares of families are gathered around the term school. Other terms associated with the school are among the terms spoken by families. Due to the use of the phrase "first," it can be assumed that the second most essential aspect is the first dosage of the vaccine. It is hoped that by delving into further detail, the families' requests for the first dose of the vaccine will come to the fore. It has been noted that the age groups of children and the requirement for immunization based on their age are mentioned in all of these. It has been determined that besides vaccination, vaccine effectiveness and mask are the words that families highlight.

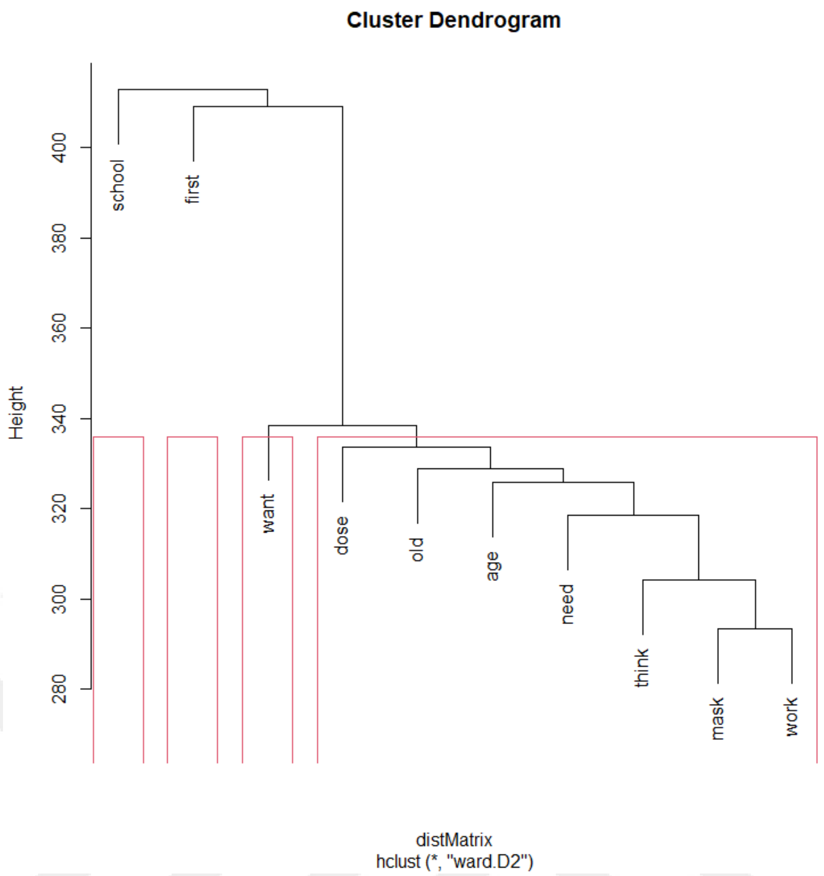

Figure 4.5: Cluster dendrogram of associated terms.

After these steps, it is time to create topic clusters for more robust results. In this direction, 4 topic clusters and the 5 most spoken terms for each topic cluster are presented in **Figure 4.6**. Then, topic clusters were tagged in tweets and saved as a CSV file for later use in sentiment analysis.

```
> terms(lda_fit, 5)
     Topic 1 Topic 2 Topic 3  Topic 4
[1,] "dose"  "first" "school" "old"
[2,] "first" "work"  "need"   "age"
[3,] "think" "want"  "first"  "want"
[4,] "age"   "dose"  "think"  "need"
[5,] "need"  "mask"  "old"    "mask"
```

Figure 4.6: 4 topic clusters and the 5 most spoken terms for each topic cluster.

Figure 4.7 was created according to beta values based on the probability of words being found in documents to improve vision and execute interpretations. When the results are analyzed one by one on a cluster basis, the main themes revealed by the topic clusters with the terms that make them up were continuously formed and interpreted;

1. The opinion of the need for first dose vaccination by age
2. First dose vaccine efficacy

3. The idea of the need for vaccination of school-aged children
4. The need for vaccinations arising from the protection of unvaccinated children only with a mask protection

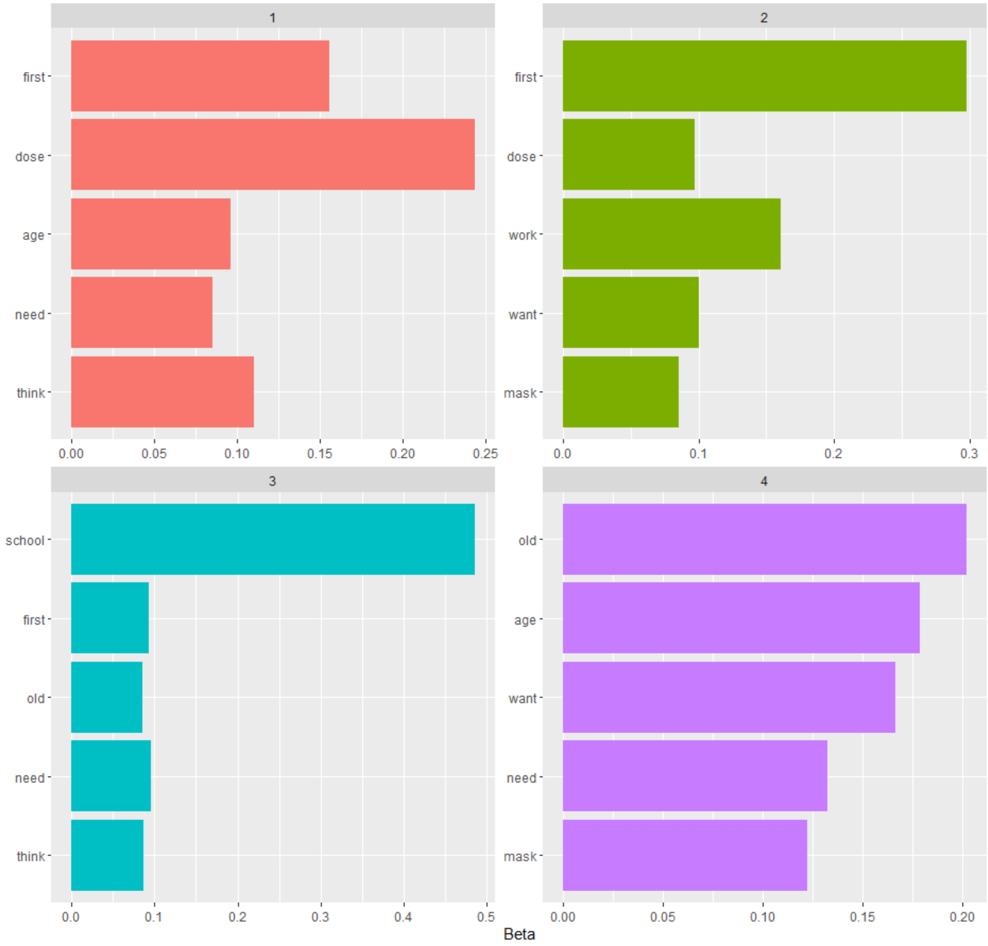

Figure 4.7: Topic clusters based on beta values.

With the determination of these 4 topics, sentiment analysis methods were used to determine the attitudes of families in getting their children vaccinated in more detail and to put forward a model. Thus, with the help of the topics tagged to each tweet in the dataset, the sentiment score was calculated for each topic cluster.

4.2. Sentiment Analysis Results

The results of the topic modeling were extracted into CSV and used to tag the tweets. The tagged tweets were organized into four different CSV files based on the topic clusters to which they belonged. The draft code, which was developed with sentiment analysis in mind, was replicated for data sets containing tweets from four different topic clusters. In this way, 4 different datasets were processed, and their results were observed using different visualizations.

4.2.1. Topic 1: The opinion of the need for first dose vaccination by age results

As a result of the topic modeling, the first topic came from the opinion of the need for first dose vaccination by age. This has come to the fore because of the idea of considering the age status of families in administering the first dose of vaccines to children, which is already on the agenda in some countries and has not yet been on the agenda in some others. Sentiment analysis was carried out to observe the attitudes and feelings of the parents on the subject and to develop interpretations about the topic cluster. The first cluster of topics was drawn into the R software program and the emotion score was calculated in the first step. The emotion scores calculated for the first 10 rows are shared in **Figure 4.8**. As seen in the figure, the anger and fear scores for the first tweet were high compared to other ones, and accordingly, the negativity of the tweet was higher.

```
> head (d,10)
   anger anticipation disgust fear joy sadness surprise trust negative positive
1      2            0       0    2   0       0        0     0        3        1
2      0            0       0    0   0       0        0     0        0        0
3      0            0       0    0   0       0        0     0        0        2
4      1            3       2    3   0       3        1     0        4        2
5      0            2       0    0   2       0        1     3        0        4
6      1            1       0    3   0       1        0     2        2        3
7      1            2       1    3   1       3        2     2        4        4
8      2            1       1    4   0       2        0     0        5        0
9      1            0       0    3   0       0        0     3        3        3
10     0            4       0    3   0       1        0     1        2        2
```

Figure 4.8: Emotion scores for first 10 rows in Topic 1.

After this step, a plot was drawn with the help of sentiment scores of all tweets belonging to topic 1. Thus, the general sentiment of topic 1 was determined and interpreted. As seen in **Figure 4.9**, it has been observed that people have trust in the opinion of the need for first dose vaccination by age, besides, they have anticipation, but despite all these, fear also outweighs. As a result, it has been determined that people expect vaccination according to their kids' age despite their fears. The fact that the results came in this way also suggests that the reason for this fear may be that their children are unvaccinated. In addition, it has been observed that the feelings of trust towards the first dose of vaccine applications to their children based on age have been formed.

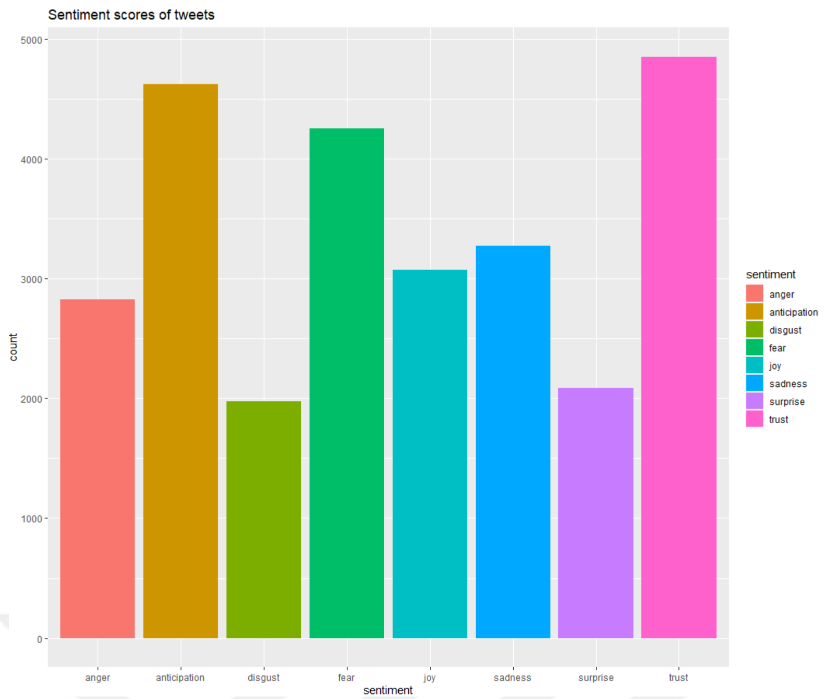

Figure 4.9: Sentiment scores of topic 1's tweets.

In addition, when observed in **Figure 4.10**, it is seen that these interpretations are supported. It is a critical fact that the emotions of trust, anticipation, and fear account for more than 15% of the dataset that contains tweets.

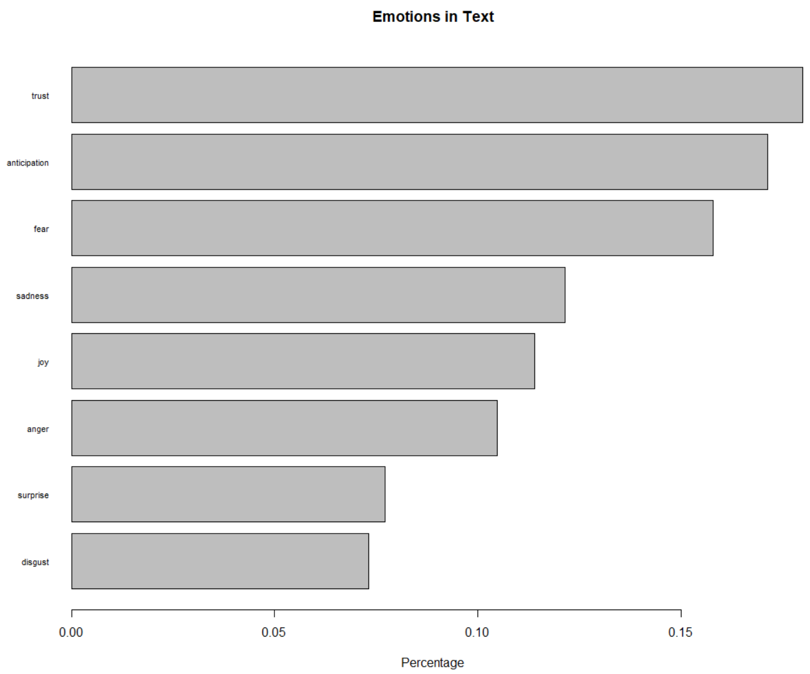

Figure 4.10: Emotions in text for topic 1.

In order to improve the comments made about the tweets, the output in **Figure 4.11** was created with the help of the code written to understand which side of the data is closer in terms of positivity and negativity. When the output is examined, it is seen that the first tweet contains more negativity, the second tweet has neutral emotions, and the third tweet is closer to positive emotions.

```
> d[1:10,]
   anger anticipation disgust fear joy sadness surprise trust negative positive score
1      2            0       0    2   0       0        0     0        3        1    -2
2      0            0       0    0   0       0        0     0        0        0     0
3      0            0       0    0   0       0        0     0        0        2     2
4      1            3       2    3   0       3        1     0        4        2    -2
5      0            2       0    0   2       0        1     3        0        4     4
6      1            1       0    3   0       1        0     2        2        3     1
7      1            2       1    3   1       3        2     2        4        4     0
8      2            1       1    4   0       2        0     0        5        0    -5
9      1            0       0    3   0       0        0     3        3        3     0
10     0            4       0    3   0       1        0     1        2        2     0
```

Figure 4.11: Positivity and negativity scores of tweets for topic 1.

Afterward, the tweets were examined in general and the relevant emotion scores were calculated to be able to comment in a general context (see **Figure 4.12**). In addition, with the help of different visualizations such as **Figure 4.13** and **Figure 4.14**, it was observed more clearly which emotion came to the fore. As a result, it was decided to assist the positive perspectives to outweigh the negative perspectives in the results, that is, the families have a positive expectation for vaccination to be started based on age.

```
anger anticipation  disgust   fear    joy  sadness  surprise   trust  negative  positive  score
 2825         4622     1974   4257   3075     3277      2084    4855      6759      7740    981
```

Figure 4.12: Tweet scores in terms of different emotions for topic 1.

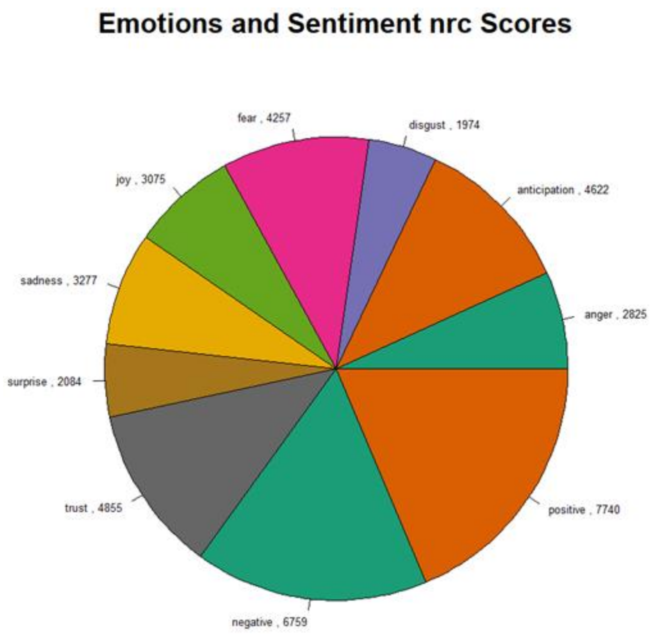

Figure 4.13: Circle chart visualization of sentiments for topic 1.

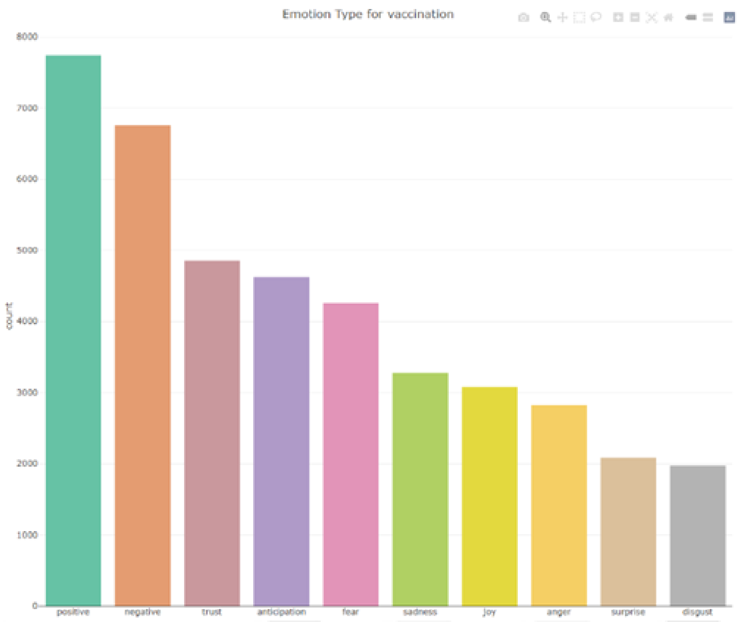

Figure 4.14: Emotion type classification on a bar plot for topic 1.

In line with the evaluations made, the information about which words the prominent emotions are gathered around is of great importance for the analysis. The word cloud created in this direction and presented in **Figure 4.15** shapes the analysis. When the terms of the sense of trust that stand out in the analysis are examined, it is seen that the trust that families have towards the people who support their children during the pandemic process, such as the government, doctors, teachers, and nurses, comes to the fore, especially for the first dose of vaccine. In addition, when the words belonging to the feeling of fear are observed, it is understood that the fear that occurs is actually due to the risks of the vaccine since it will be the first one that children experienced and that the children who have the vaccine are afraid of observing the side effects. Finally, it is seen that families are willing to start making vaccines by age for children based on the sentiment of anticipation, one of the prominent emotions in the direction of the first topic cluster.

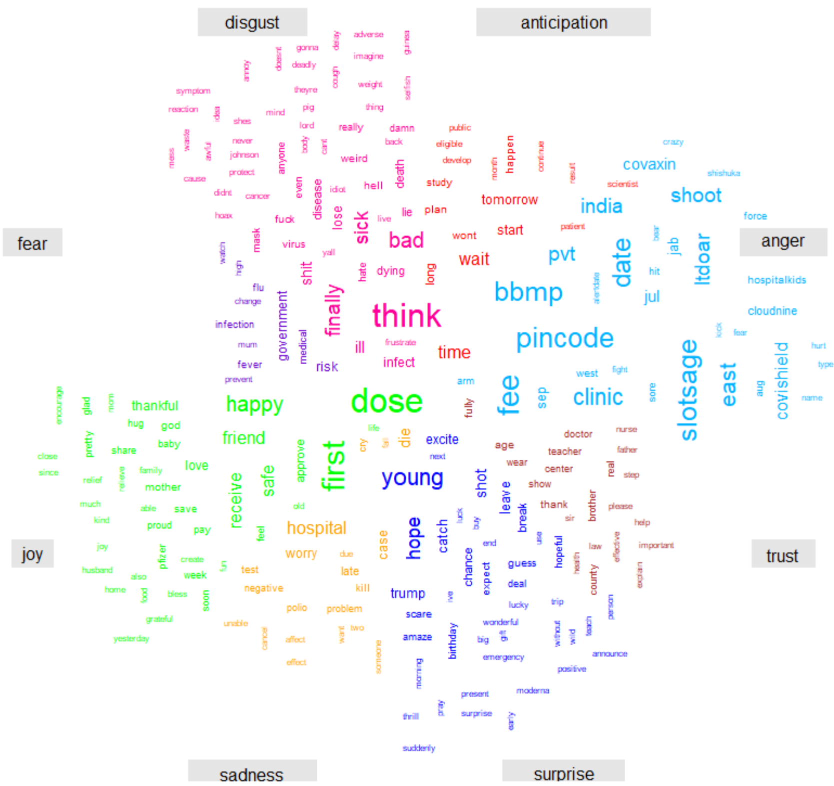

Figure 4.15: Word cloud created in terms of emotions for topic 1.

As a result, when the tweets belonging to the first topic cluster were analyzed by sentiment analysis methods, it was seen that the idea of the age-related need for vaccination of families was identified with trust in the government. In addition, it was observed that the sentiment of anticipation was also emphasized. It was commented that in this case, families were willing to vaccinate their children and that they trusted the suggestions of people such as doctors, nurses, and teachers who supported their children during the pandemic process. Although positive emotions were dominant in the results, it was observed in the first set of subjects that there was an intense fear of risks and side effects in families as a result of vaccination.

4.2.2. Topic 2: First dose vaccine efficacy age results

The second cluster of issues concerned the efficacy of first-dose vaccines. This also shows that; Families observed the results of the first dose of vaccines around the world and shared on Twitter about the effectiveness of the first vaccination dosage to be administered to their children. Their feelings about the effectiveness of vaccination are one of the most crucial elements to shape their ideas and attitudes about getting their children vaccinated. For each remaining topic cluster, the steps performed in topic 1 were followed, and interpretations were made as the outputs were obtained. In this direction, firstly, tweets related to topic 2 were drawn into the program. Thus, the counting of tweets based on emotion was started. For example, when **Figure 4.16** is examined, anger, fear, surprise, and negative emotions predominate for the first line, which is a sign of negative interpretations about vaccine efficacy in this section.

```
> head (d,10)
   anger anticipation disgust fear joy sadness surprise trust negative positive
1      2            0       0    2   0       0        2     0        2        0
2      0            2       0    2   0       1        0     2        1        1
3      2            1       1    4   0       1        1     0        5        0
4      0            1       0    1   1       1        0     1        1        1
5      1            0       2    2   0       2        0     0        4        1
6      0            0       0    1   0       1        0     0        1        1
7      1            2       1    1   1       1        1     3        1        2
8      1            2       1    1   1       1        2     1        1        3
9      1            0       0    1   0       1        0     1        2        1
10     1            0       1    1   0       1        1     1        1        2
```

Figure 4.16: Emotion scores for first 10 rows in Topic 2.

Considering all the tweets in more detail, it was observed that there was a similar situation in Topic 2 as well as in Topic 1. When **Figure 4.17** and **Figure 4.18** are examined, it has been determined that the feeling of trust and anticipation is dominant in tweets, and the feeling of fear also comes to the fore. In addition, when the dominance percentages of the emotions in the data set are examined, it is seen that the feelings of trust, anticipation, and fear, which are again prominent, find a great place for them.

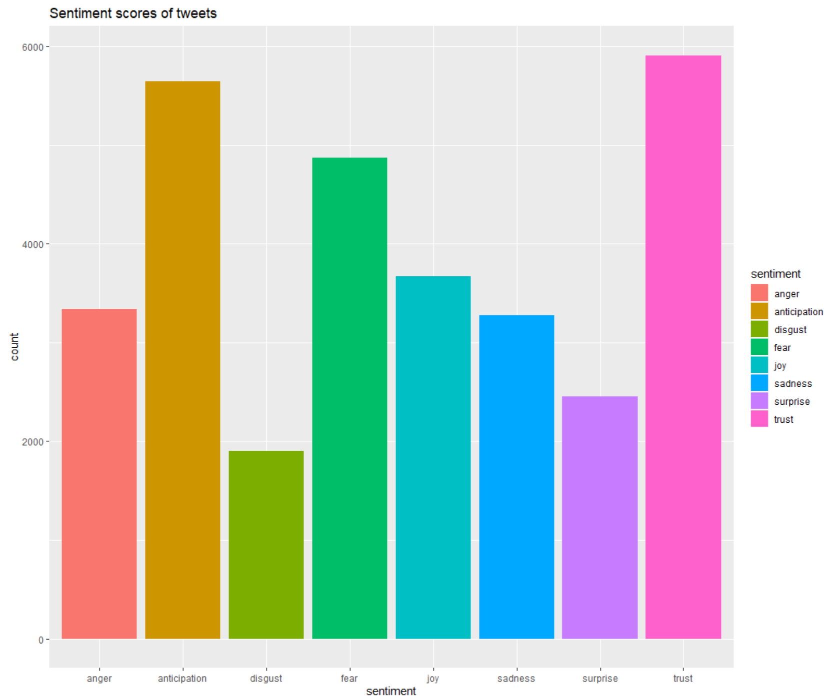

Figure 4.17: Sentiment scores of topic 2's tweets.

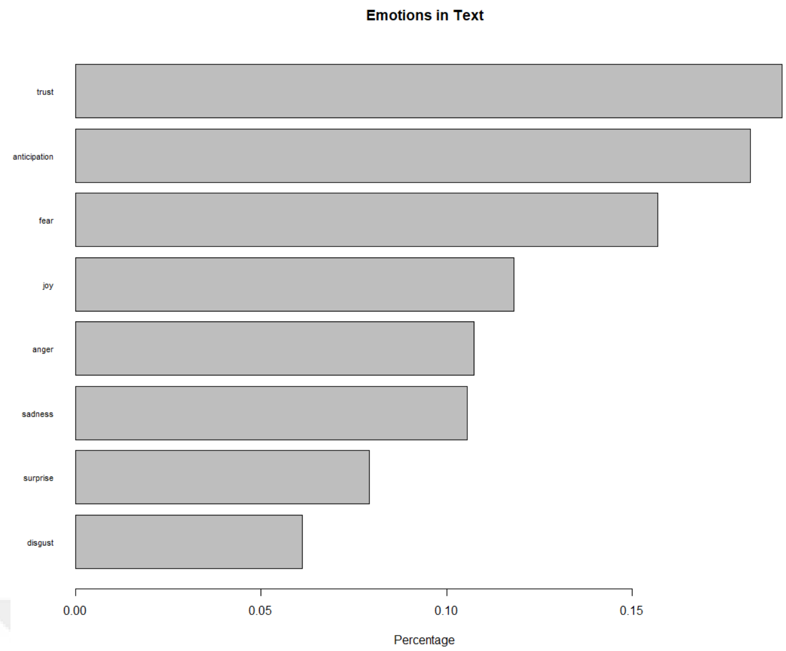

Figure 4.18: Emotions in text for topic 2.

Following the observation of the main emotions, the output in **Figure 4.19** was reviewed to see if the tweets were close to emotional negativity or positivity, and an overall score was produced, as shown in **Figure 4.20**. Taking all of this into account, it was determined that the majority of the topic cluster 2 participants expressed pleasant emotions.

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive	score
1	2	0	0	2	0	0	2	0	2	0	-2
2	0	2	0	2	0	1	0	2	1	1	0
3	2	1	1	4	0	1	1	0	5	0	-5
4	0	1	0	1	1	1	0	1	1	1	0
5	1	0	2	2	0	2	0	0	4	1	-3
6	0	0	0	1	0	1	0	0	1	1	0
7	1	2	1	1	1	1	1	3	1	2	1
8	1	2	1	1	1	1	2	1	1	3	2
9	1	0	0	1	0	1	0	1	2	1	-1
10	1	0	1	1	0	1	1	1	1	2	1

Figure 4.19: Positivity and negativity scores of tweets for topic 2.

anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive	score
3336	5646	1899	4869	3668	3278	2457	5911	6831	9163	2332

Figure 4.20: Tweet scores in terms of different emotions for topic 2.

With the support of the different visualizations used in **Figure 4.21** and **Figure 4.22**, it is seen that positive emotions are well ahead, and this is supported by the feelings of trust and anticipation. Many of the negative emotions are caused by the fear of families about having their children vaccinated.

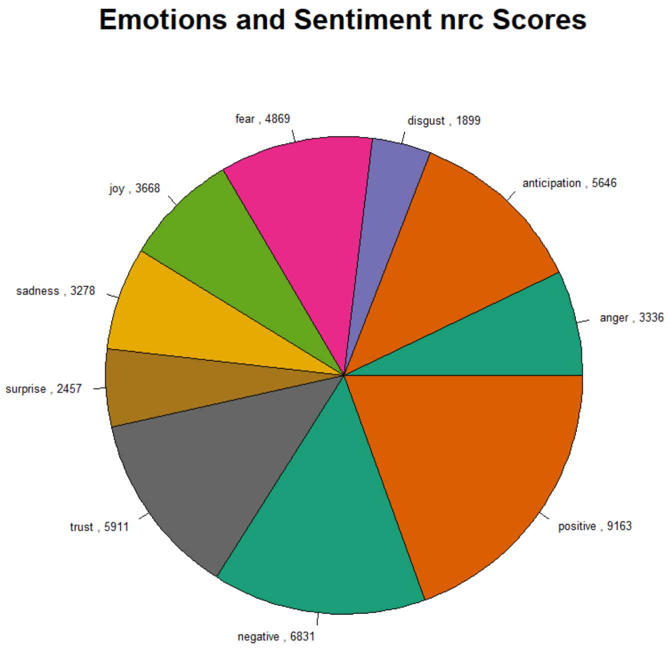

Figure 4.21: Circle chart visualization of sentiments for topic 2.

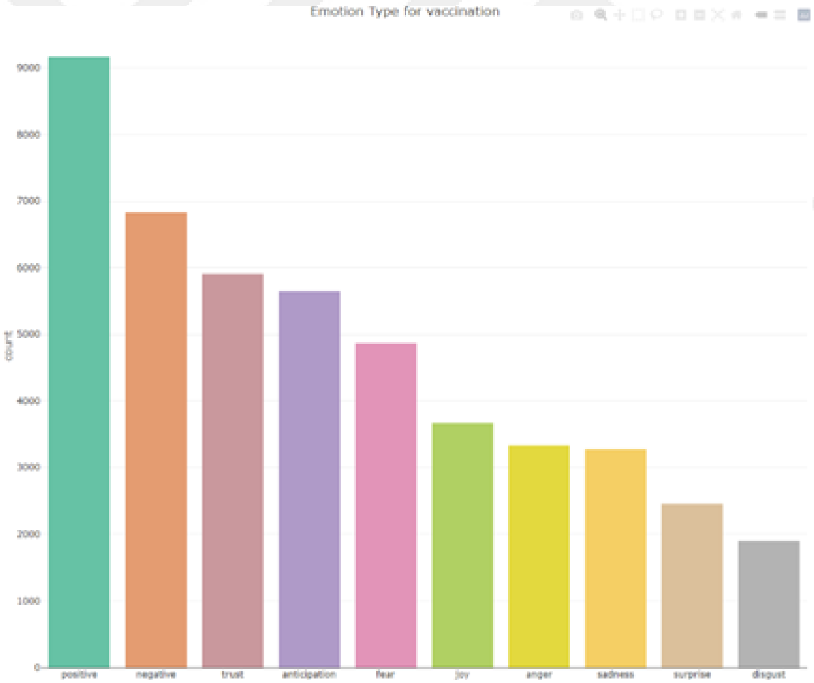

Figure 4.22: Emotion type classification on a bar plot for topic 2.

With the creation of the word cloud in **Figure 4.23**, the terms around which the prominent emotions were gathered were observed, and all the results were reinforced. As seen in the word cloud, the feeling of trust, which is the most dominant emotion in tweets, is gathered around words such as president, teacher, announce, world, and nurse. This shows the trust of families in the coronavirus vaccine release, announced by presidents, also, suggested by trustworthy persons such as teachers and health workers in terms of the efficacy situation of the vaccine. In addition, when the feeling of anticipation was examined, it was commented that new vaccines were expected and that the parents thought that the existing vaccines should be developed and given to their children. Finally, another prominent emotion, the feeling of fear, was examined and it was seen that the fear and anxiety of the families were also related to the side effects that the vaccine may bring, namely indirect risks since they don't know about the vaccine efficacy.

As a result, it has been observed that families generally have a positive opinion about the effectiveness of the first dose of vaccine, the feelings of trust, expectation, and fear come to the fore just like in the first topic cluster. It has been seen that the sense of trust in vaccine effectiveness is gathered around the coronavirus vaccine developed and announced to the whole world by presidents, also, it has been examined that the expectations for the development of new vaccines for more effective results have intensified. Despite all this sense of trust, it has been observed that when the subject is their children, families experience uneasiness and fear of possible risks.

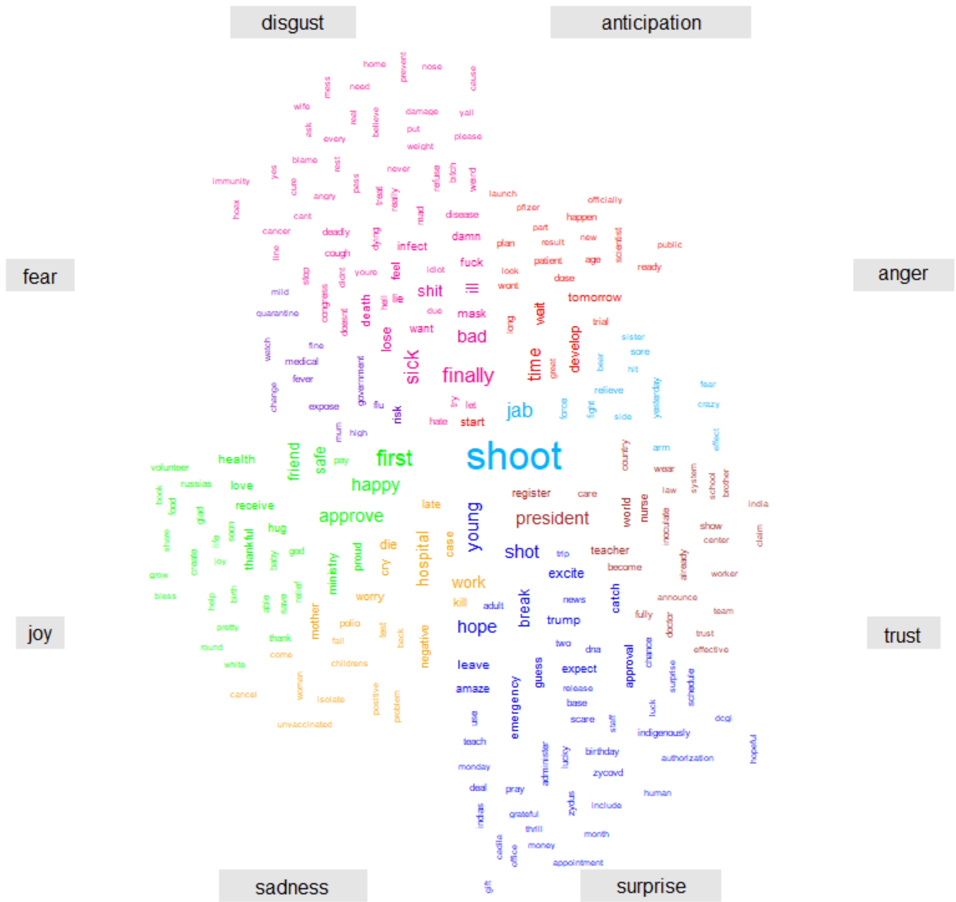

Figure 4.23: Word cloud created in terms of emotions for topic 2.

4.2.3. Topic 3: The idea of the need for vaccination of school-aged children results

The third cluster of topics came as the idea of the need for vaccination of school-aged children. The most important reason for this is undoubtedly the opening of schools that have been online for a long time and turning to face-to-face education. Families have started to argue about vaccination because children who have not been in a social environment for a long time enter a crowded environment with their peers. In this context, as seen in **Figure 4.24**, the emotional intensities of the tweets were observed first. For example, only the feeling of trust was found in the first tweet, which indicated that it was towards the more neutral side.

```
> head (d,10)
   anger anticipation disgust fear joy sadness surprise trust negative positive
1      0            0       0    0   0       0        0     1        0        0
2      0            1       0    1   2       3        0     3        3        5
3      0            0       0    0   0       0        0     1        0        0
4      1            1       0    2   1       3        1     3        5        1
5      0            0       0    1   0       0        0     2        0        3
6      0            1       0    0   0       0        0     1        0        0
7      0            0       0    0   2       1        1     2        1        3
8      0            1       0    0   0       0        0     2        0        3
9      0            2       0    1   1       0        0     2        1        3
10     0            0       0    0   0       0        0     1        0        0
```

Figure 4.24: Emotion scores for first 10 rows in Topic 3.

The distribution of the emotional intensity of the topic cluster, in general, is given in **Figure 4.25** and **Figure 4.26** as a percentage. Unlike other topic clusters, it was determined that the feeling of trust left other emotions in the background. When the percentage distribution was examined, it was observed that more than 30% of the whole emotion score was formed by the feeling of trust. In addition, the emotions of sadness, anticipation, and fear were also prominent compared to other emotional states.

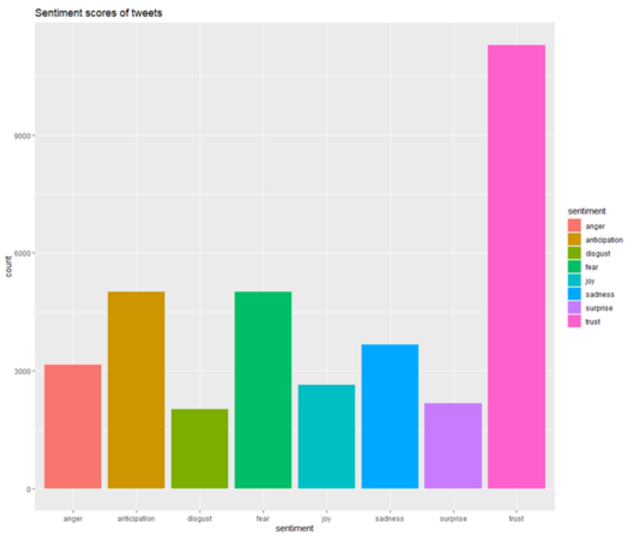

Figure 4.25: Sentiment scores of topic 3's tweets.

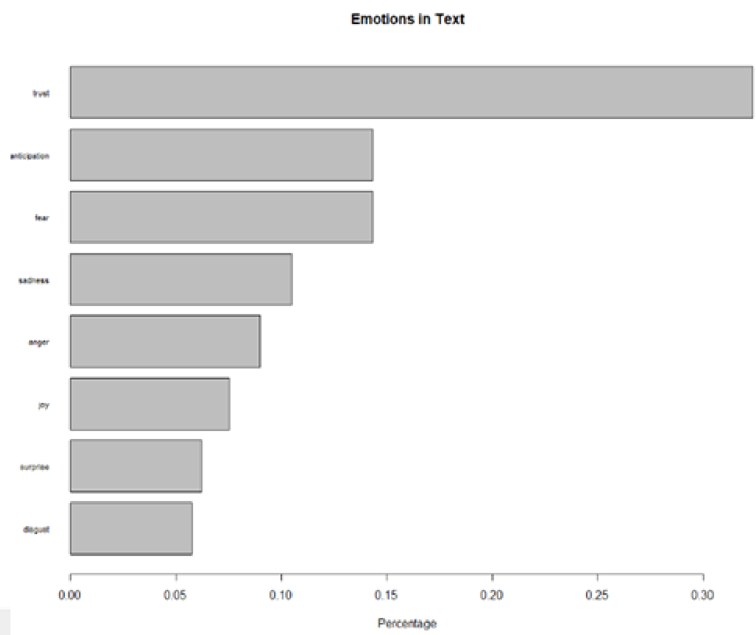

Figure 4.26: Emotions in text for topic 3.

After observing the dominance of the feeling of trust in the topic cluster, support was obtained from **Figure 4.27** to determine the proximity of the tweets to the positive, negative, or neutral side to be able to comment. When the figure was examined, it was found that the first tweet had neutral feelings, the second tweet had positivity, and the fourth tweet was full of highly negative thoughts.

```
> d[1:10,]
   anger anticipation disgust fear joy sadness surprise trust negative positive score
1      0            0       0    0   0       0        0     1        0        0     0
2      0            1       0    1   2       3        0     3        3        5     2
3      0            0       0    0   0       0        0     1        0        0     0
4      1            1       0    2   1       3        1     3        5        1    -4
5      0            0       0    1   0       0        0     2        0        3     3
6      0            1       0    0   0       0        0     1        0        0     0
7      0            0       0    0   2       1        1     2        1        3     2
8      0            1       0    0   0       0        0     2        0        3     3
9      0            2       0    1   1       0        0     2        1        3     2
10     0            0       0    0   0       0        0     1        0        0     0
```

Figure 4.27: Positivity and negativity scores of tweets for topic 3.

After this step, the tweets were examined in a general context and it was observed that the 3rd data set was close to the positive side, just like in other data sets (see **Figure 4.28** and **Figure 4.29**). In addition, when analyzed in **Figure 4.30**, it was determined that the feeling of trust was more prominent than even the positivity and negativeness scores. In other words, it has been observed that families trust the strategies of policymakers to vaccinate their school-age children, and they carry this to an important point.

```
> print(tweetscore)
   anger anticipation      disgust         fear          joy      sadness     surprise        trust     negative     positive        score
    3148         5006         2011         5001         2634         3668         2167        11298         7895         8773          878
```

Figure 4.28: Tweet scores in terms of different emotions for topic 3.

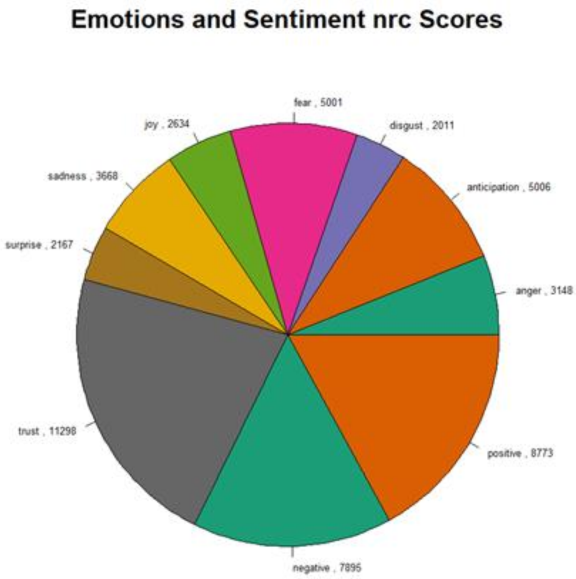

Figure 4.29: Circle chart visualization of sentiments for topic 3.

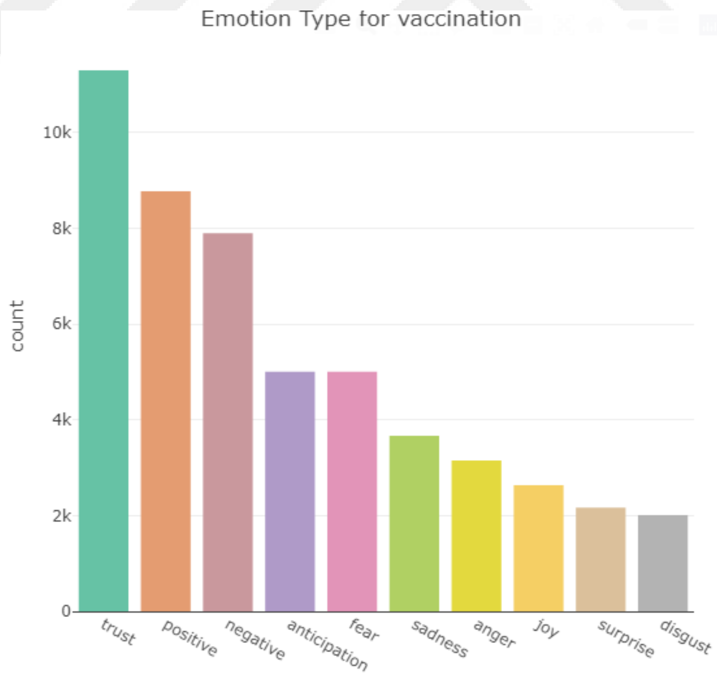

Figure 4.30: Emotion type classification on a bar plot for topic 3.

Finally, an emotion-based word cloud was created to carry out detailed interpretations and to finalize the sentiment analysis for the 3rd topic cluster (see **Figure 4.31**). The terms around the feeling of trust, which is the most prominent emotion, are gathered to explain the reason for the emotional intensity of the topic cluster. When the word cloud was examined, it was determined that the families wanted to impose an obligation for vaccination, it was preferred for the individuals to be at the school to be vaccinated, and in this way, the feeling of trust came to the fore. In addition, it has been observed that the feeling of expectation is shaped around the development of vaccines and giving them to children, while the emotion of fear is caused by the danger-risk situation that the vaccine will bring.

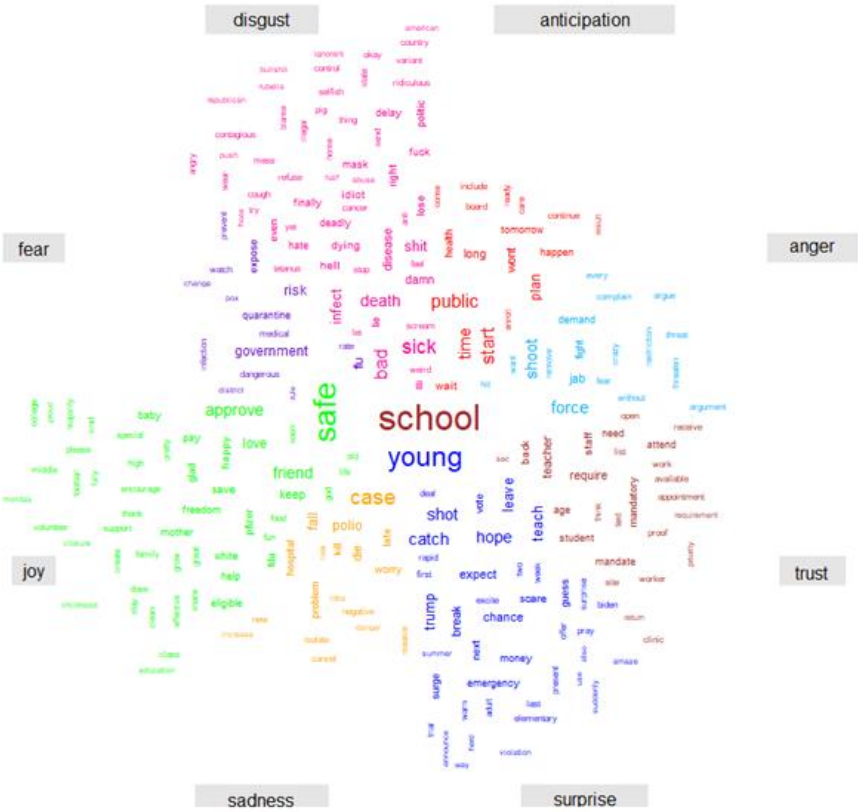

Figure 4.31: Word cloud created in terms of emotions for topic 3.

As a result, when the opinions of families about the need to vaccinate their school-age children are examined based on emotion, it has been found that generally positive emotions come to the fore and the feeling of trust predominates. It has been commented that families are willing to open schools by ensuring the environment in which school-age children will be found, that is, by making vaccination compulsory for everyone who works and is a student at school. It has also been observed that parents are willing to develop new vaccines and administer the first dose of vaccine to their children, but they are afraid of the risks. They saw the way to protect their children, who are not yet vaccinated, as vaccinating the staff at the school.

4.2.4. Topic 4: The need for vaccinations arising from the protection of unvaccinated children only with a mask protection results

The dataset, which includes the ideas of the need for vaccination arising from the protection of unvaccinated children only with masks, which is the 4th and last topic set, was also transferred to the software program and processed. This set of topics also emphasizes that children whose immunity is not fully formed and who are open to all diseases can only be protected with a mask against a serious epidemic that is the coronavirus. For this reason, the critical need for the coronavirus vaccine to be given to children comes to the fore. Additionally, the fact that this topic has more tweets compared to other topic clusters has turned attention to its results. In this context, when the emotion scores were calculated based on the tweets (see **Figure 4.32**), it was seen by examining the first tweet as an example; the first tweet contains many different emotional aspects.

```
> head (d,10)
```

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	1	1	1	1	0	1	0	1	2	0
2	2	0	2	3	1	2	0	3	3	4
3	0	1	0	1	0	0	0	1	2	2
4	0	2	0	0	0	0	0	0	0	0
5	1	1	1	1	0	0	0	0	3	0
6	1	0	1	2	0	2	0	2	3	1
7	0	0	0	0	1	0	0	1	0	2
8	0	1	1	0	3	0	0	2	0	4
9	0	0	0	0	0	0	0	0	1	0
10	0	1	0	0	1	0	1	0	0	1

Figure 4.32: Emotion scores for first 10 rows in Topic 4.

Then, with the help of emotion scores calculated one by one, the prominent emotions in the data set were determined with the help of different visualizations. In this context, when **Figure 4.33** is examined, it is seen that the feeling of trust comes to the fore, similar to other topics, the feeling of fear outweighs a little more, and the anticipation of the families is intense. Likewise, when analyzed in **Figure 4.34**, it was seen that these emotions constituted more than 15% of the whole data set as a percentage.

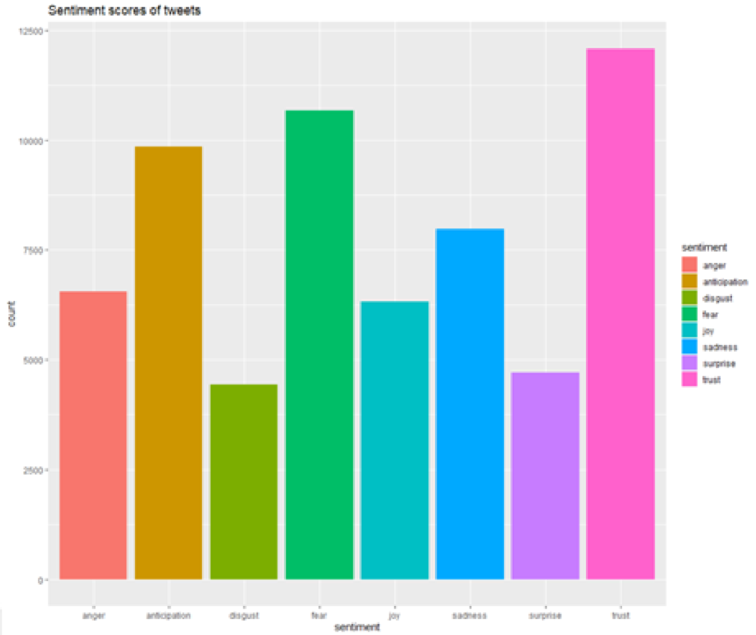

Figure 4.33: Sentiment scores of topic 4's tweets.

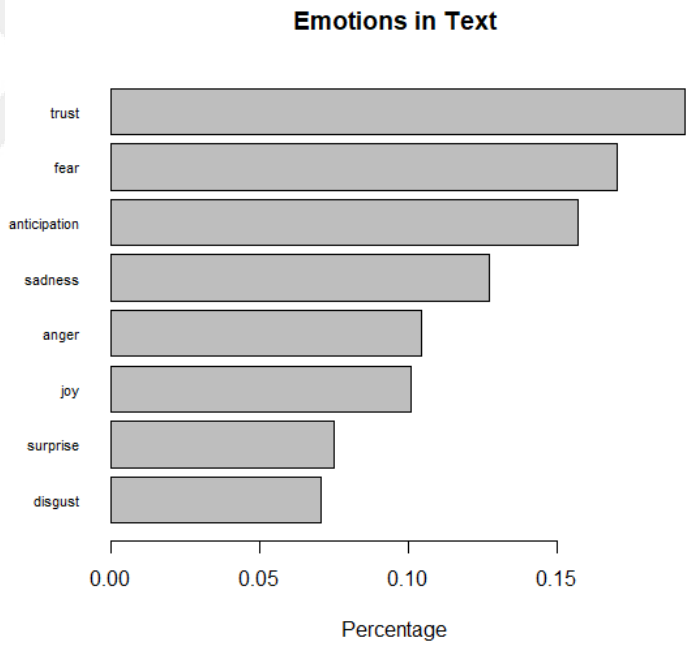

Figure 4.34: Emotions in text for topic 4.

In addition, the output in **Figure 4.35** was created to find out whether the tweets are closer to positivity or negativity. It has been observed that the tweets in the first 10 rows generally have a negative perspective.

```
> d[1:10,]
   anger anticipation disgust fear joy sadness surprise trust negative positive score
1      1            1       1    1   0       1        0     1        2        0    -2
2      2            0       2    3   1       2        0     3        3        4     1
3      0            1       0    1   0       0        0     1        2        2     0
4      0            2       0    0   0       0        0     0        0        0     0
5      1            1       1    1   0       0        0     0        3        0    -3
6      1            0       1    2   0       2        0     2        3        1    -2
7      0            0       0    0   1       0        0     1        0        2     2
8      0            1       1    0   3       0        0     2        0        4     4
9      0            0       0    0   0       0        0     0        1        0    -1
10     0            1       0    0   1       0        1     0        0        1     1
```

Figure 4.35: Positivity and negativity scores of tweets for topic 4.

In order to observe the general emotional intensity of the data set, the output in **Figure 4.36** was created and interpretations were carried out with the support of the visualizations in **Figure 4.37** and **Figure 4.38**. First of all, although the closeness of negative and positive emotions was striking, it was determined that positive emotions were the dominant side in the data set. Besides, it was commented that all other emotions had scores close to each other, but trust, fear, and anticipation were prominent when looking at the first three.

```
> print(tweetscore)
       anger anticipation      disgust        fear         joy     sadness    surprise       trust    negative    positive       score
        6552         9846         4441       10666        6329        7976        4709       12087       16744       17374         630
```

Figure 4.36: Tweet scores in terms of different emotions for topic 4.

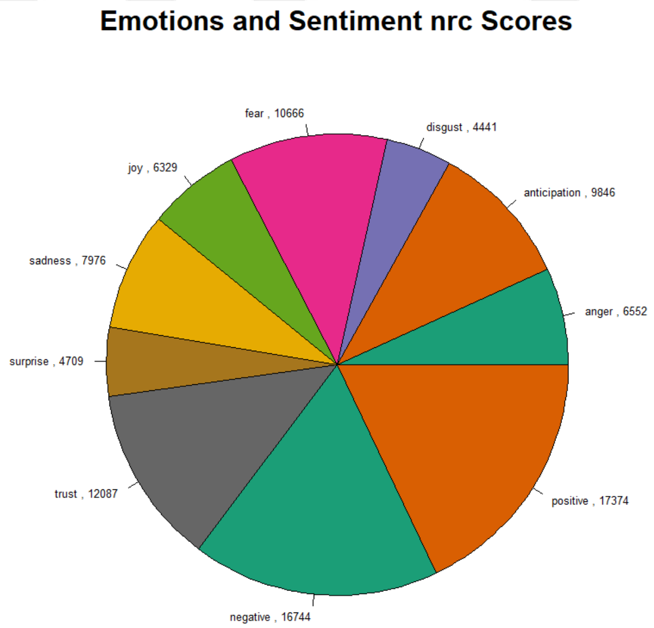

Figure 4.37: Circle chart visualization of sentiments for topic 4.

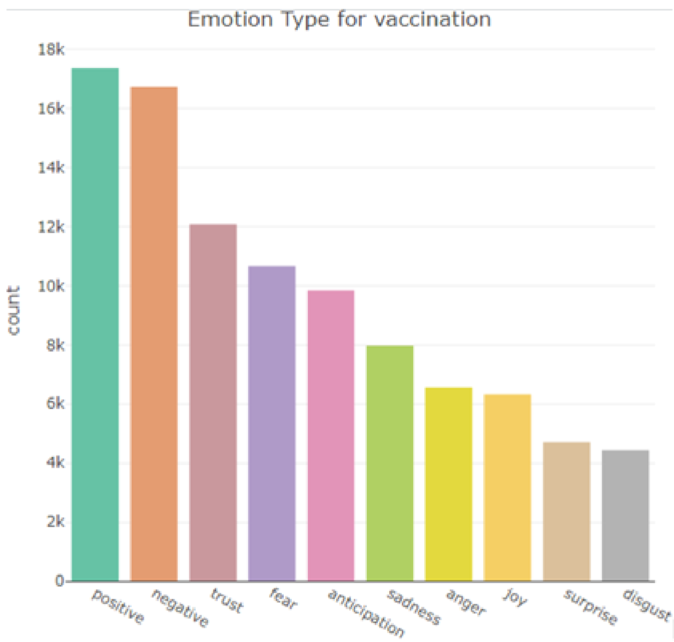

Figure 4.38: Emotion type classification on a bar plot for topic 4.

In order to observe the reason for the intense emotions, a word cloud was created, and the terms included in the foreground emotions were examined (see **Figure 4.39**). In this context, it has been observed that terms such as age, mask, Pfizer, and effectiveness constitute the feeling of trust, which is the emotion with the highest percentage. Thus, while it is thought that people trust the Pfizer vaccine to be made based on age, it has been commented that they also believe in the protection of wearing a mask. Secondly, the feeling of fear was examined and the dominance of words such as infection, quarantine, high, and government was determined. It was thought that people's fear was for their children to become infected after the quarantine and they argued that the government should do something about it. Thirdly, the sense of anticipation was examined, and people's expectations for a new vaccine or the development of an existing vaccine came to the fore, just like in other topic clusters.

In the 4th topic cluster, it was observed that the positive emotions, although close to the negative, were dominant, while the feelings of trust, fear and anticipation were predominant. It has been determined that families trust the protection of the mask, but at the same time expect a new vaccine. However, when it comes to their children, it has been understood that they are afraid of possible infectious situations and the consequences that may come with the vaccine.

As a result, it was seen that positive emotions predominated in the 4 topic clusters and three emotions came to the fore;

• Trust

• Anticipation

• Fear

It is thought that the most important reason for this is that children are at a very sensitive point for their families. Families respect the strategic measures taken by the government or decision-makers and trust their protection. But with the relaxation of the measures taken, they are worried about their children who have no immunity to the coronavirus. This brings the vaccine debate to the fore. With the return to normal social life and the opening of schools, families have come to expect the coronavirus vaccines to be given to their children. In this way, they follow the developments all over the world, also, they are anticipation to develop existing vaccines and conducting new vaccine trials. However, when it comes to their children, they are afraid of the risks that the vaccine will bring, and they are afraid of the side effects of this coronavirus vaccine.

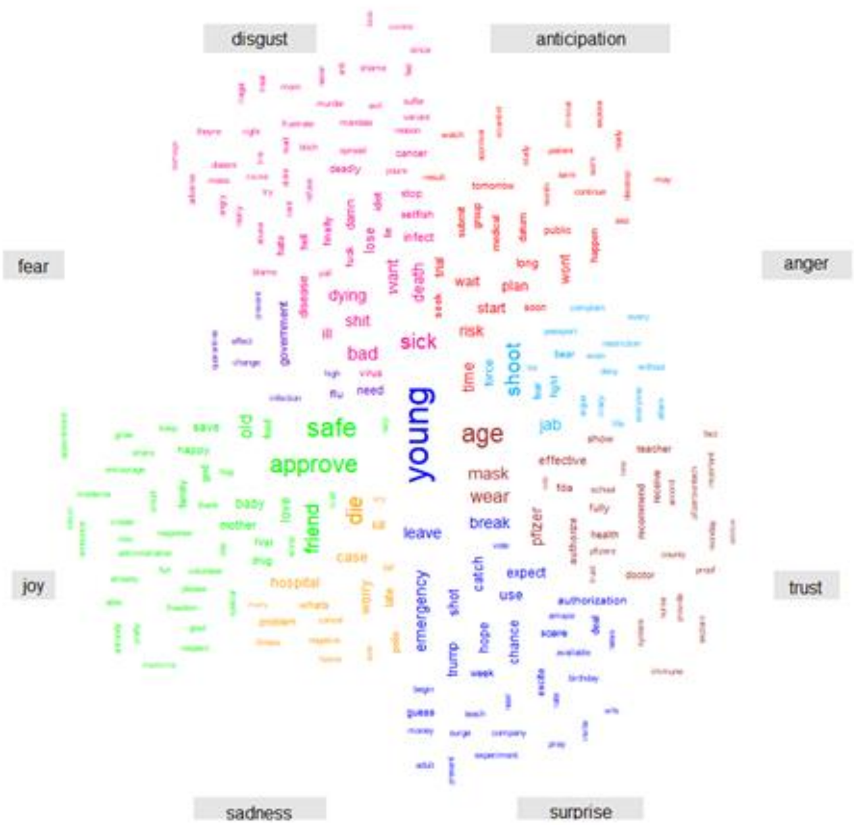

Figure 4.39: Word cloud created in terms of emotions for topic 4.

In addition to this, it can be said that the opening of schools also causes concern for families. It is thought that the number of cases will increase as the schools open and return to face-to-face education for children who are at home during the pandemic period and continue their education online. This makes families think about the best way to protect children. Many sharing made by families with children on Twitter brought the issues related to the vaccination of children to a critical point during the opening of schools.

CHAPTER 5

5. CONCLUSIONS AND FUTURE WORK

At the end of 2019, the coronavirus epidemic emerged in Wuhan, China, spread all over the world, and became a common problem for everyone. In order to mitigate the course of the epidemic, policymakers started to apply measures and precautions. These strategies significantly restricted the lives of parents and children, causing them to leave the normal life routines they were accustomed to. As a result of this situation, children have faced many negative psychological effects due to the disruption of their normal routines.

Society accepted that the most important way to cope with COVID-19 is the vaccines developed against this disease. However, many studies argue that the majority of the world's population should be vaccinated in order to gain herd immunity. While the discussions on vaccines continue, families have created a big question mark in the minds of how the vaccination process of children will proceed. The reactions of families have echoed through popular social media platforms. Parents not only shared their attitudes with people but also followed the agenda, also, other people's thoughts in this way.

In this study, it is aimed to analyze the attitudes of families about getting their children vaccinated and determine the reasons behind these attitudes. It has been observed that researchers have been examining the reactions of society against COVID-19 since the beginning of the epidemic. The authors focused on small target audiences in their studies, preferred the traditional method of survey, and tried to determine the general opinion. In contrast to this situation, some researchers focus on new methods. There is a particular trend toward NLP methods, which have grown in popularity in recent years. To the best of our knowledge, no study has been found in the literature that reveals parents' attitudes toward vaccination of their children using a single model that combines topic modeling and sentiment analysis techniques. Several social media platforms have become indispensable communication tools for people in the course of the epidemic. One of these platforms is Twitter, a social media tool where people can express their thoughts transparently. For this reason, Twitter social media platform was preferred in this study, and the related data were created as links by using an advanced search with the support of keywords. In order to make this data workable, the Octoparse tool, which is frequently used by researchers and has a simple structure, was used. Thus, a data set containing tweets of families about children's vaccination was created. With the help of the created dataset and RStudio, the targeted model has been put forward. First, data preprocessing was performed. Then, topic clusters were created with the support of LDA, and the created clusters were processed with sentiment analysis, thus, the views of families on getting their children vaccinated were revealed with the support of the established model.

As a result of the study, four topic clusters were determined which were the opinion of the need for first dose vaccination by age, first dose vaccine efficacy, the idea of the need for vaccination of school-aged children, and the need for vaccinations arising from the protection of unvaccinated children only with mask protection according to frequently used words. By the processing of the determined topic clusters with sentiment analysis, it was determined that positive emotions were predominant, and 3 emotions, namely trust, expectation, and fear, came to the fore.

By looking at the outputs of the model presented, it has been observed that families trust the measures taken to alleviate the epidemic, but when these measures are relaxed, they begin to worry about their children. Parents have started to anticipate vaccinations for their children from different perspectives with the start of returning to normal life. However, it has been determined that the most important point that creates a question mark in their minds is the risk and side effects of the vaccine. With the help of the word cloud created in topic modeling, it is thought that the most important reason that fuels these discussions is the school that has started face-to-face education. Families who think that the number of cases will increase, especially based on age, have brought vaccination to the fore. As a result, it has been determined that families trust the states and their announcements about getting their children vaccinated, they anticipate new vaccines to be developed, but they are also afraid of the risks that the vaccine will bring to their children.

During the study period, some limitations occurred in the data collection phase. Since this study cannot focus on each language separately, it has been focused only on tweets sent in English, which is a universal language, and this limited the target audience. In addition, the sentiment libraries of RStudio were used in the sentiment analysis phase, and emotion scores were calculated. In future studies, it is thought that libraries can be created according to the data set to be used in the sentiment analysis phase to achieve clearer and desired results with higher accuracy.

The increasing diversity and importance of NLP studies in recent years to catch up with the process of the coronavirus have been demonstrated by a detailed study. It is thought that the study will shed light on the professionals who develop strategies in the field of health during the process development stages and the researchers in their study.

BIBLIOGRAPHY

[1] J. G. Shim, K. H. Ryu, S. H. Lee, E. A. Cho, Y. J. Lee, and J. H. Ahn, 'Text mining approaches to analyze public sentiment changes regarding covid-19 vaccines on social media in korea', *Int. J. Environ. Res. Public Health*, vol. 18, no. 12, 2021, doi: 10.3390/ijerph18126549.

[2] K. A. Fisher, S. J. Bloomstone, J. Walder, S. Crawford, H. Fouayzi, and K. M. Mazor, 'Attitudes Toward a Potential SARS-CoV-2 Vaccine : A Survey of U.S. Adults', *Ann. Intern. Med.*, vol. 173, no. 12, pp. 964–973, 2020, doi: 10.7326/M20-3569.

[3] 'COVID Live Update: 244,536,555 Cases and 4,966,160 Deaths from the Coronavirus - Worldometer'. https://www.worldometers.info/coronavirus/ (accessed Oct. 25, 2021).

[4] 'WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data'. https://covid19.who.int/ (accessed Oct. 25, 2021).

[5] A. Hussain *et al.*, 'Artificial intelligence-enabled analysis of public attitudes on facebook and twitter toward COVID-19 vaccines in the United Kingdom and the United States: Observational study', *J. Med. Internet Res.*, vol. 23, no. 4, pp. 1–10, 2021, doi: 10.2196/26627.

[6] L. Li *et al.*, 'Characterizing the Propagation of Situational Information in Social Media during COVID-19 Epidemic: A Case Study on Weibo', *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 2, pp. 556–562, 2020, doi: 10.1109/TCSS.2020.2980007.

[7] S. Ballı, 'Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods Serkan', no. January, 2020.

[8] S. Connor, 'The relationship between media and covid 19', no. December 2020, 2021.

[9] M. Català *et al.*, 'Monitoring and Analysis of COVID-19 Pandemic: The Need for an Empirical Approach', *Front. Public Heal.*, vol. 9, no. July, pp. 1–9, 2021, doi: 10.3389/fpubh.2021.633123.

[10] S. Sharfuddin, 'The world after Covid-19', *Round Table*, vol. 109, no. 3, pp. 247–257, 2020, doi: 10.1080/00358533.2020.1760498.

[11] S. J. Barnes, 'Information management research and practice in the post-COVID-19 world', *Int. J. Inf. Manage.*, vol. 55, no. June, p. 102175, 2020, doi: 10.1016/j.ijinfomgt.2020.102175.

[12] Y. Zhang, L. Li, Y. Jiang, and B. Huang, 'Analysis of COVID-19 prevention and control effects based on the seitrd dynamic model and Wuhan epidemic statistics', *Int. J. Environ. Res. Public Health*, vol. 17, no. 24, pp. 1–16, 2020, doi: 10.3390/ijerph17249309.

[13] W. Tan *et al.*, 'Is returning to work during the COVID-19 pandemic stressful? A study on immediate mental health status and psychoneuroimmunity prevention measures of Chinese workforce', *Brain. Behav. Immun.*, vol. 87, no. April, pp. 84–

92, 2020, doi: 10.1016/j.bbi.2020.04.055.

[14] V. M. Lamarche, ‘Socially connected and COVID-19 prepared: The influence of sociorelational safety on perceived importance of COVID-19 precautions and trust in government responses’, *Soc. Psychol. Bull.*, vol. 15, no. 4, 2020, doi: 10.32872/spb.4409.

[15] ‘Coronavirus (COVID-19) Cases - Statistics and Research - Our World in Data’. https://ourworldindata.org/covid-cases (accessed Oct. 29, 2021).

[16] ‘School closures during the COVID-19 pandemic, Oct 27, 2021’. https://ourworldindata.org/grapher/school-closures-covid?time=latest (accessed Oct. 30, 2021).

[17] S. W. Patrick *et al.*, ‘Well-being of Parents and Children During the COVID-19 Pandemic: A National Survey’, *Pediatrics*, vol. 146, no. 4, p. e2020016824, 2020, doi: 10.1542/peds.2020-016824.

[18] S. Abuhammad, ‘Parents’ knowledge and attitude towards COVID-19 in children: A Jordanian Study’, *Int. J. Clin. Pract.*, vol. 75, no. 2, pp. 1–6, 2021, doi: 10.1111/ijcp.13671.

[19] N. Singhal, D. D. McMillan, R. Savich, D. Matovelo, D. Santorino, and B. D. Kamath-Rayne, ‘Development and impact of helping babies breathe educational methodology’, *Pediatrics*, vol. 146, no. 4, 2020, doi: 10.1542/peds.2020-007294.

[20] ‘COVID-19 confirmed cases and deaths - UNICEF DATA’. https://data.unicef.org/resources/covid-19-confirmed-cases-and-deaths-dashboard/ (accessed Oct. 30, 2021).

[21] P. G. Szilagyi *et al.*, ‘Parents’ Intentions and Perceptions About COVID-19 Vaccination for Their Children: Results From a National Survey’, *Pediatrics*, vol. 148, no. 4, p. e2021052335, 2021, doi: 10.1542/peds.2021-052335.

[22] A. Chandarakesan, S. Muruhan, and R. R. A. Sayanam, ‘Morin Inhibiting Photocarcinogenesis by Targeting Ultraviolet-B-Induced Oxidative Stress and Inflammatory Cytokines Expression in Swiss Albino Mice’, *Int. J. Nutr. Pharmacol. Neurol. Dis*, vol. 8, pp. 41–46, 2018, doi: 10.4103/ijnpnd.ijnpnd.

[23] M. Spinelli, F. Lionetti, M. Pastore, and M. Fasolo, ‘Parents’ Stress and Children’s Psychological Problems in Families Facing the COVID-19 Outbreak in Italy’, *Front. Psychol.*, vol. 11, no. January, pp. 1–7, 2020, doi: 10.3389/fpsyg.2020.01713.

[24] A. Garbe, U. ogurlu, N. Logan, and P. Cook, ‘Parents’ Experiences with Remote Education during COVID-19 School Closures’, *Am. J. Qual. Res.*, vol. 4, no. 3, pp. 45–65, 2020, doi: 10.29333/ajqr/8471.

[25] T. Parczewska, ‘Difficult situations and ways of coping with them in the experiences of parents homeschooling their children during the COVID-19 pandemic in Poland’, *Educ. 3-13*, vol. 49, no. 7, pp. 889–900, 2021, doi: 10.1080/03004279.2020.1812689.

[26] M. Drouin, B. T. McDaniel, J. Pater, and T. Toscos, ‘How Parents and Their Children Used Social Media and Technology at the Beginning of the COVID-19 Pandemic and Associations with Anxiety’, *Cyberpsychology, Behav. Soc. Netw.*,

vol. 23, no. 11, pp. 727–736, 2020, doi: 10.1089/cyber.2020.0284.

[27] ‘How Many People Use Social Media in 2021? (65+ Statistics)’. https://backlinko.com/social-media-users#social-media-usage-stats (accessed Oct. 31, 2021).

[28] N. A. Obi-Ani, C. Anikwenze, and M. C. Isiani, ‘Social media and the Covid-19 pandemic: Observations from Nigeria’, *Cogent Arts Humanit.*, vol. 7, no. 1, 2020, doi: 10.1080/23311983.2020.1799483.

[29] S. Kemp, ‘Digital 2021: Global Overview Report — DataReportal – Global Digital Insights’. https://datareportal.com/reports/digital-2021-global-overview-report (accessed Nov. 01, 2021).

[30] D. Chaffey, ‘Global social media statistics research summary [updated 2021]’. https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/ (accessed Oct. 31, 2021).

[31] D. Valdez, M. ten Thij, K. Bathina, L. A. Rutter, and J. Bollen, ‘Social Media Insights Into US Mental Health During the COVID-19 Pandemic: Longitudinal Analysis of Twitter Data’, *J. Med. Internet Res.*, vol. 22, no. 12, Dec. 2020, doi: 10.2196/21418.

[32] M. Jeyanathan, S. Afkhami, F. Smaill, M. S. Miller, B. D. Lichty, and Z. Xing, ‘Immunological considerations for COVID-19 vaccine strategies’, *Nat. Rev. Immunol.*, vol. 20, no. 10, pp. 615–632, 2020, doi: 10.1038/s41577-020-00434-6.

[33] A. A. Malik, S. A. M. McFadden, J. Elharake, and S. B. Omer, ‘Determinants of COVID-19 vaccine acceptance in the US’, *EClinicalMedicine*, vol. 26, p. 100495, 2020, doi: 10.1016/j.eclinm.2020.100495.

[34] ‘Coronavirus (COVID-19) Vaccinations - Statistics and Research - Our World in Data’. https://ourworldindata.org/covid-vaccinations (accessed Nov. 01, 2021).

[35] L. Corey, J. R. Mascola, A. S. Fauci, and F. S. Collins, ‘A strategic approach to COVID-19 vaccine R&D’, *Science (80-.).*, vol. 368, no. 6494, pp. 948–950, May 2020, doi: 10.1126/SCIENCE.ABC5312.

[36] B. S. Graham, ‘Rapid COVID-19 vaccine development’, *Science (80-.).*, vol. 368, no. 6494, pp. 945–946, May 2020, doi: 10.1126/SCIENCE.ABB8923.

[37] ‘Coronavirus disease (COVID-19): Vaccines’. https://www.who.int/news-room/q-a-detail/coronavirus-disease-(covid-19)-vaccines?adgroupsurvey=%7Badgroupsurvey%7D&gclid=CjwKCAjwoP6LBhBlEiwAvCcthDyeJQurK5950RPiyA5it7lki2VWeQmPCCtNmnCA3XJorQ8VGAhfDBoChBMQAvD_BwE (accessed Nov. 01, 2021).

[38] W. Y. S. Chou and A. Budenz, ‘Considering Emotion in COVID-19 Vaccine Communication: Addressing Vaccine Hesitancy and Fostering Vaccine Confidence’, *Health Commun.*, vol. 35, no. 14, pp. 1718–1722, 2020, doi: 10.1080/10410236.2020.1838096.

[39] A. A. Dror *et al.*, ‘Vaccine hesitancy: the next challenge in the fight against COVID-19’, *Eur. J. Epidemiol. 2020 358*, vol. 35, no. 8, pp. 775–779, Aug. 2020, doi: 10.1007/S10654-020-00671-Y.

[40] S. Machingaidze and C. S. Wiysonge, ‘Understanding COVID-19 vaccine hesitancy’, *Nat. Med. 2021 278*, vol. 27, no. 8, pp. 1338–1339, Jul. 2021, doi: 10.1038/s41591-021-01459-7.

[41] M. Sallam, ‘COVID-19 Vaccine Hesitancy Worldwide: A Concise Systematic Review of Vaccine Acceptance Rates’, *Vaccines 2021, Vol. 9, Page 160*, vol. 9, no. 2, p. 160, Feb. 2021, doi: 10.3390/VACCINES9020160.

[42] M. Sallam, ‘Covid-19 vaccine hesitancy worldwide: A concise systematic review of vaccine acceptance rates’, *Vaccines*, vol. 9, no. 2, pp. 1–15, 2021, doi: 10.3390/vaccines9020160.

[43] J. Murphy *et al.*, ‘Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom’, *Nat. Commun. 2021 121*, vol. 12, no. 1, pp. 1–15, Jan. 2021, doi: 10.1038/s41467-020-20226-9.

[44] ‘COVID-19 aşılarının ve tedavilerinin takibi | McKinsey’. https://www.mckinsey.com/industries/life-sciences/our-insights/on-pins-and-needles-will-covid-19-vaccines-save-the-world (accessed Nov. 04, 2021).

[45] P. G. Szilagyi *et al.*, ‘Parents’ intentions and perceptions about COVID-19 vaccination for their children: Results from a national survey’, *Pediatrics*, vol. 148, no. 4, 2021, doi: 10.1542/peds.2021-052335.

[46] K. C. Zhang *et al.*, ‘Parental acceptability of COVID-19 vaccination for children under the age of 18 years: Cross-sectional online survey’, *JMIR Pediatr. Parent.*, vol. 3, no. 2, pp. 1–13, 2020, doi: 10.2196/24827.

[47] M. Yigit, A. Ozkaya-Parlakay, and E. Senel, ‘Evaluation of COVID-19 Vaccine Refusal in Parents’, *Pediatr. Infect. Dis. J.*, vol. 40, no. 4, pp. E134–E136, 2021, doi: 10.1097/INF.0000000000003042.

[48] S. Bell, R. Clarke, S. Mounier-Jack, J. L. Walker, and P. Paterson, ‘Parents’ and guardians’ views on the acceptability of a future COVID-19 vaccine: A multi-methods study in England’, *Vaccine*, vol. 38, no. 49, pp. 7789–7798, 2020, doi: 10.1016/j.vaccine.2020.10.027.

[49] L. Singh *et al.*, ‘A first look at COVID-19 information and misinformation sharing on Twitter.’, *ArXiv*, 2020, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/32550244%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7280879.

[50] K. F. Ali, S. Whitebridge, M. H. Jamal, M. Alsafy, and S. L. Atkin, ‘Perceptions, knowledge, and behaviors related to COVID-19 among social media users: Cross-sectional study’, *J. Med. Internet Res.*, vol. 22, no. 9, pp. 1–10, 2020, doi: 10.2196/19913.

[51] W. Cuiyan *et al.*, ‘Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China’, *Int. J. Environ. Res. Public Health*, vol. 17, no. 5, pp. 1–25, 2020.

[52] A. S. Raamkumar, S. G. Tan, and H. L. Wee, ‘Measuring the Outreach Efforts of Public Health Authorities and the Public Response on Facebook during the COVID-19 Pandemic in Early 2020: Cross-Country Comparison’, *J. Med. Internet*

Res., vol. 22, no. 5, 2020, doi: 10.2196/19334.

[53] Fitzpatrick K, Harris C, and Drawve G, 'Fear of COVID-19 and the Mental Health Consequences in America. Psychological Trauma: Theory, Research, Practice, and Policy', vol. 12, pp. 17–21, 2020, [Online]. Available: https://content.apa.org/fulltext/2020-38568-001.pdf.

[54] H. W. Park, S. Park, and M. Chong, 'Conversations and medical news frames on twitter: Infodemiological study on COVID-19 in South Korea', *J. Med. Internet Res.*, vol. 22, no. 5, 2020, doi: 10.2196/18897.

[55] W. Zhen, 'Parental Perception about the Effects of Covid-19 Pandemic on Young Children By', 2020.

[56] S. Yeasmin *et al.*, 'Impact of COVID-19 pandemic on the mental health of children in Bangladesh: A cross-sectional study', *Child. Youth Serv. Rev.*, vol. 117, no. July, p. 105277, 2020, doi: 10.1016/j.childyouth.2020.105277.

[57] N. L. Alfieri *et al.*, 'Parental COVID-19 vaccine hesitancy for children : vulnerability in an urban hotspot', pp. 1–9, 2021.

[58] D. Scannell *et al.*, 'COVID-19 Vaccine Discourse on Twitter: A Content Analysis of Persuasion Techniques, Sentiment and Mis/Disinformation', *J. Health Commun.*, vol. 26, no. 7, pp. 443–459, 2021, doi: 10.1080/10810730.2021.1955050.

[59] J. V. Lazarus *et al.*, 'A global survey of potential acceptance of a COVID-19 vaccine', *Nat. Med.*, vol. 27, no. 2, pp. 225–228, 2021, doi: 10.1038/s41591-020-1124-9.

[60] J. Khubchandani, S. Sharma, J. H. Price, M. J. Wiblishauser, M. Sharma, and F. J. Webb, 'COVID-19 Vaccination Hesitancy in the United States: A Rapid National Assessment', *J. Community Health*, vol. 46, no. 2, pp. 270–277, 2021, doi: 10.1007/s10900-020-00958-x.

[61] S. Loomba, A. de Figueiredo, S. J. Piatek, K. de Graaf, and H. J. Larson, 'Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA', *Nat. Hum. Behav.*, vol. 5, no. 3, pp. 337–348, 2021, doi: 10.1038/s41562-021-01056-1.

[62] J. K. Ward *et al.*, 'The French public's attitudes to a future COVID-19 vaccine: The politicization of a public health issue', *Soc. Sci. Med.*, vol. 265, pp. 2016–2021, 2020, doi: 10.1016/j.socscimed.2020.113414.

[63] Z. Hou *et al.*, 'Assessing covid-19 vaccine hesitancy, confidence, and public engagement:a global social listening study', *J. Med. Internet Res.*, vol. 23, no. 6, pp. 1–11, 2021, doi: 10.2196/27632.

[64] P. L. Reiter, M. L. Pennell, and M. L. Katz, 'Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information', *Vaccine*, vol. 38, no. January, pp. 6500–6507, 2020, [Online]. Available: https://doi.org/10.1016/j.vaccine.2020.08.043.

[65] L. C. Jiang, T. H. Chu, and M. Sun, 'Characterization of Vaccine Tweets During

the Early Stage of the COVID-19 Outbreak in the United States: Topic Modeling Analysis', *JMIR Infodemiology*, vol. 1, no. 1, p. e25636, 2021, doi: 10.2196/25636.

[66] M. M. Davis, J. S. Zickafoose, A. E. Halvorson, and S. Patrick, 'Parents' Likelihood to Vaccinate Their Children and Themselves Against COVID-19', *medRxiv*, p. 2020.11.10.20228759, 2020, [Online]. Available: http://medrxiv.org/content/early/2020/11/13/2020.11.10.20228759.abstract.

[67] R. D. Goldman *et al.*, 'Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information ', no. January, 2020.

[68] M. Skjefte *et al.*, 'COVID-19 vaccine acceptance among pregnant women and mothers of young children: results of a survey in 16 countries', *Eur. J. Epidemiol.*, vol. 36, no. 2, pp. 197–211, 2021, doi: 10.1007/s10654-021-00728-6.

[69] M. Yılmaz and M. K. Sahin, 'Parents' willingness and attitudes concerning the COVID-19 vaccine: A cross-sectional study', *Int. J. Clin. Pract.*, vol. 75, no. 9, pp. 1–11, 2021, doi: 10.1111/ijcp.14364.

[70] M. C. Jenkins and M. A. Moreno, 'Vaccination Discussion among Parents on Social Media: A Content Analysis of Comments on Parenting Blogs', *J. Health Commun.*, vol. 25, no. 3, pp. 232–242, 2020, doi: 10.1080/10810730.2020.1737761.

[71] I. J. B. Young, S. Luz, and N. Lone, 'A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis', *Int. J. Med. Inform.*, vol. 132, p. 103971, 2019, doi: 10.1016/j.ijmedinf.2019.103971.

[72] E. Pons, L. M. M. Braun, M. G. M. Hunink, and J. A. Kors, 'Natural language processing in radiology: A systematic review', *Radiology*, vol. 279, no. 2, pp. 329–343, 2016, doi: 10.1148/radiol.16142770.

[73] M. Masum *et al.*, 'Actionable Knowledge Extraction Framework for COVID-19', *Proc. - 2020 IEEE Int. Conf. Big Data, Big Data 2020*, pp. 4036–4041, 2020, doi: 10.1109/BigData50022.2020.9378398.

[74] A. Z. Klein, A. Magge, K. O'Connor, J. I. F. Amaro, D. Weissenbacher, and G. G. Hernandez, 'Toward Using Twitter for Tracking COVID-19:A Natural Language Processing Pipeline and Exploratory Data Set', *J. Med. Internet Res.*, vol. 23, no. 1, 2021, doi: 10.2196/25314.

[75] J. Ayoub, X. J. Yang, and F. Zhou, 'Combat COVID-19 infodemic using explainable natural language processing models', *Inf. Process. Manag.*, vol. 58, no. 4, 2021, doi: 10.1016/j.ipm.2021.102569.

[76] P. Bose, S. Roy, and P. Ghosh, 'A Comparative NLP-Based Study on the Current Trends and Future Directions in COVID-19 Research', *IEEE Access*, vol. 9, pp. 78341–78355, 2021, doi: 10.1109/ACCESS.2021.3082108.

[77] L. Tang *et al.*, 'Texas public agencies' tweets and public engagement during the COVID-19 pandemic: Natural language processing approach', *JMIR Public Heal. Surveill.*, vol. 7, no. 4, 2021, doi: 10.2196/26720.

[78] N. Sadman, N. Anjum, K. Datta Gupta, and M. A. Parvez Mahmud, 'Understanding the Pandemic through Mining Covid News Using Natural Language Processing', *2021 IEEE 11th Annu. Comput. Commun. Work. Conf. CCWC 2021*, pp. 362–367, 2021, doi: 10.1109/CCWC51732.2021.9376002.

[79] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, 'COVID-19 public sentiment insights and machine learning for tweets classification', *Inf.*, vol. 11, no. 6, pp. 1–22, 2020, doi: 10.3390/info11060314.

[80] X. Sun *et al.*, 'Analysis on Action Tracking Reports of COVID-19 Informs Control Strategies and Vaccine Delivery in Post-Pandemic Era', *medRxiv*, p. 2021.04.08.21254953, 2021, [Online]. Available: http://medrxiv.org/content/early/2021/04/25/2021.04.08.21254953.abstract.

[81] Z. T. Osakwe, I. Ikhapoh, B. K. Arora, and O. M. Bubu, 'Identifying public concerns and reactions during the COVID-19 pandemic on Twitter: A text-mining analysis', *Public Health Nurs.*, vol. 38, no. 2, pp. 145–151, 2021, doi: 10.1111/phn.12843.

[82] N. Gozzi *et al.*, 'Collective Response to Media Coverage of the COVID-19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis', *J. Med. Internet Res.*, vol. 22, no. 10, pp. 1–19, 2020, doi: 10.2196/21597.

[83] J. C. Lyu and G. K. Luli, 'Understanding the public discussion about the centers for disease control and prevention during the COVID-19 pandemic using twitter data: Text mining analysis study', *J. Med. Internet Res.*, vol. 23, no. 2, 2021, doi: 10.2196/25108.

[84] A. Agarwal, P. Salehundam, S. Padhee, W. L. Romine, and T. Banerjee, 'Leveraging Natural Language Processing to Mine Issues on Twitter during the COVID-19 Pandemic', *Proc. - 2020 IEEE Int. Conf. Big Data, Big Data 2020*, pp. 886–891, 2020, doi: 10.1109/BigData50022.2020.9378028.

[85] C. Doogan, W. Buntine, H. Linger, and S. Brunt, 'Public perceptions and attitudes toward covid-19 nonpharmaceutical interventions across six countries: A topic modeling analysis of twitter data', *J. Med. Internet Res.*, vol. 22, no. 9, pp. 1–11, 2020, doi: 10.2196/21419.

[86] Y. Liu, C. Whitfield, T. Zhang, A. Hauser, T. Reynolds, and M. Anwar, 'Monitoring COVID-19 pandemic through the lens of social media using natural language processing and machine learning', *Heal. Inf. Sci. Syst.*, vol. 9, no. 1, pp. 1–16, 2021, doi: 10.1007/s13755-021-00158-4.

[87] S. Liu, J. Li, and J. Liu, 'Leveraging transfer learning to analyze opinions, attitudes, and behavioral intentions toward COVID-19 vaccines: Social media content and temporal analysis', *J. Med. Internet Res.*, vol. 23, no. 8, 2021, doi: 10.2196/30251.

[88] Q. Liu *et al.*, 'Health communication through news media during the early stage of the covid-19 outbreak in China: Digital topic modeling approach', *J. Med. Internet Res.*, vol. 22, no. 4, 2020, doi: 10.2196/19118.

[89] X. Han, J. Wang, M. Zhang, and X. Wang, 'Using social media to mine and analyze public opinion related to COVID-19 in China', *Int. J. Environ. Res. Public Health*, vol. 17, no. 8, 2020, doi: 10.3390/ijerph17082788.

[90] M. H. Tsai and Y. Wang, 'Analyzing twitter data to evaluate people's attitudes towards public health policies and events in the era of covid-19', *Int. J. Environ. Res. Public Health*, vol. 18, no. 12, 2021, doi: 10.3390/ijerph18126272.

[91] A. Ahmed, C. Argho, D. Suben, and K. Saha, 'Sentiment Analysis of COVID-19 Vaccination from Survey Responses in Bangladesh', pp. 1–15, 2021.

[92] M. O. Lwin *et al.*, 'Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends', *JMIR Public Heal. Surveill.*, vol. 6, no. 2, pp. 1–4, 2020, doi: 10.2196/19447.

[93] B. P. Pokharel, 'Twitter Sentiment Analysis During Covid-19 Outbreak in Nepal', *SSRN Electron. J.*, no. March, pp. 1–9, 2020, doi: 10.2139/ssrn.3624719.

[94] A. D. Dubey, 'The resurgence of cyber racism during the COVID-19 pandemic and its aftereffects: Analysis of sentiments and emotions in tweets', *JMIR Public Heal. Surveill.*, vol. 6, no. 4, pp. 1–7, 2020, doi: 10.2196/19833.

[95] M. Singh, A. K. Jakhar, and S. Pandey, 'Sentiment analysis on the impact of coronavirus in social life using the BERT model', *Soc. Netw. Anal. Min.*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1007/s13278-021-00737-z.

[96] S. Boon-Itt and Y. Skunkan, 'Public perception of the COVID-19 pandemic on twitter: Sentiment analysis and topic modeling study', *JMIR Public Heal. Surveill.*, vol. 6, no. 4, pp. 1–17, 2020, doi: 10.2196/21978.

[97] H. Yin, S. Yang, and J. Li, 'Detecting Topic and Sentiment Dynamics Due to COVID-19 Pandemic Using Social Media', *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12447 LNAI, no. June, pp. 610–623, 2020, doi: 10.1007/978-3-030-65390-3_46.

[98] A. Al-Rawi, K. Grepin, X. Li, R. Morgan, C. Wenham, and J. Smith, 'Investigating Public Discourses Around Gender and COVID-19: a Social Media Analysis of Twitter Data', *J. Healthc. Informatics Res.*, vol. 5, no. 3, pp. 249–269, 2021, doi: 10.1007/s41666-021-00102-x.

[99] L. Chen, H. Lyu, T. Yang, Y. Wang, and J. Luo, 'In the Eyes of the Beholder: Analyzing Social Media Use of Neutral and Controversial Terms for COVID-19', pp. 3–6, 2020, [Online]. Available: http://arxiv.org/abs/2004.10225.

[100] M. Hung *et al.*, 'Social network analysis of COVID-19 sentiments: Application of artificial intelligence', *J. Med. Internet Res.*, vol. 22, no. 8, pp. 1–13, 2020, doi: 10.2196/22590.

[101] A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hai, and Z. Shah, 'Top concerns of tweeters during the COVID-19 pandemic: A surveillance study', *J. Med. Internet Res.*, vol. 22, no. 4, pp. 1–9, 2020, doi: 10.2196/19016.

[102] A. Ebadi, P. Xi, S. Tremblay, B. Spencer, R. Pall, and A. Wong, 'Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing', *Scientometrics*, vol. 126, no. 1, pp. 725–739, 2021, doi: 10.1007/s11192-020-03744-7.

[103] O. Oyebode *et al.*, 'COVID-19 Pandemic: Identifying Key Issues using Social Media and Natural Language Processing', 2020, [Online]. Available: http://arxiv.org/abs/2008.10022.

[104] I. Li, Y. Li, T. Li, S. Alvarez-Napagao, D. Garcia-Gasulla, and T. Suzumura, ‘What are we depressed about when we talk about COVID-19: Mental health analysis on tweets using natural language processing’, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12498 LNAI, pp. 358–370, 2020, doi: 10.1007/978-3-030-63799-6_27.

[105] F. Hanschmidt and A. Kersting, ‘Emotions in Covid-19 Twitter discourse following the introduction of social contact restrictions in Central Europe’, *J. Public Heal.*, 2021, doi: 10.1007/s10389-021-01613-y.

[106] R. Raju, S. Bhandari, S. A. Mohamud, and E. N. Ceesay, ‘Transfer Learning Model for Disrupting Misinformation during a COVID-19 Pandemic’, *2021 IEEE 11th Annu. Comput. Commun. Work. Conf. CCWC 2021*, pp. 245–250, 2021, doi: 10.1109/CCWC51732.2021.9376066.

[107] J. H. Lee, ‘Understanding Public Attitudes Toward COVID-19 with Twitter’, *2021 IEEE Syst. Inf. Eng. Des. Symp. SIEDS 2021*, pp. 1–6, 2021, doi: 10.1109/SIEDS52267.2021.9483708.

[108] S. Vydra and J. Kantorowicz, ‘Tracing Policy-relevant Information in Social Media: The Case of Twitter before and during the COVID-19 Crisis’, *Stat. Polit. Policy*, vol. 12, no. 1, pp. 87–127, 2021, doi: 10.1515/spp-2020-0013.

[109] M. S. Ahmed, T. T. Aurpa, and M. M. Anwar, ‘Detecting sentiment dynamics and clusters of Twitter users for trending topics in COVID-19 pandemic’, *PLoS One*, vol. 16, no. 8 August, pp. 1–20, 2021, doi: 10.1371/journal.pone.0253300.

[110] M. Mackay, T. Colangeli, D. Gillis, J. McWhirter, and A. Papadopoulos, ‘Examining social media crisis communication during early covid-19 from public health and news media for quality, content and corresponding public sentiment’, *Int. J. Environ. Res. Public Health*, vol. 18, no. 15, 2021, doi: 10.3390/ijerph18157986.

[111] S. Fazeli, D. Zamanzadeh, A. Ovalle, T. Nguyen, G. Gee, and M. Sarrafzadeh, ‘COVID-19 and Big Data: Multi-faceted Analysis for Spatio-temporal Understanding of the Pandemic with Social Media Conversations’, 2021, [Online]. Available: http://arxiv.org/abs/2104.10807.

[112] S. Duraivel and D. R. Lavanya, ‘Understanding vaccine hesitancy with application of Latent Dirichlet Allocation to Reddit Corpora’, pp. 0–12, 2021, [Online]. Available: https://doi.org/10.21203/rs.3.rs-616664/v1.

[113] K. Thurson, ‘Parenting , Vaccines , and COVID-19 : A Machine-Learning Approach’, no. May, 2021.

[114] G. K. Sandaka and B. N. Gaekwade, ‘Sentiment Analysis and Time-series Analysis for the COVID-19 vaccine Tweets’, no. May, 2021, [Online]. Available: https://www.diva-portal.org/smash/record.jsf?pid=diva2:1576180.

[115] N. S. Sattar and S. Arifuzzaman, ‘Covid-19 vaccination awareness and aftermath: Public sentiment analysis on twitter data and vaccinated population prediction in the usa’, *Appl. Sci.*, vol. 11, no. 13, 2021, doi: 10.3390/app11136128.

[116] S. Gbashi, O. A. Adebo, W. Doorsamy, and P. B. Njobeh, ‘Systematic delineation of media polarity on COVID-19 vaccines in Africa: Computational linguistic

modeling study', *JMIR Med. Informatics*, vol. 9, no. 3, 2021, doi: 10.2196/22916.

[117] W. Wu, H. Lyu, and J. Luo, 'Characterizing Discourse about COVID-19 Vaccines: A Reddit Version of the Pandemic Story', *Heal. Data Sci.*, vol. 2021, pp. 1–11, 2021, doi: 10.34133/2021/9837856.

[118] J. C. Lyu, E. Le Han, and G. K. Luli, 'Covid-19 vaccine-related discussion on twitter: Topic modeling and sentiment analysis', *J. Med. Internet Res.*, vol. 23, no. 6, pp. 1–12, 2021, doi: 10.2196/24435.

[119] S. W. H. Kwok, S. K. Vadde, and G. Wang, 'Tweet topics and sentiments relating to COVID-19 vaccination among Australian twitter users: Machine learning analysis', *J. Med. Internet Res.*, vol. 23, no. 5, 2021, doi: 10.2196/26953.

[120] A. Karami, M. Zhu, B. Goldschmidt, H. R. Boyajieff, and M. M. Najafabadi, 'COVID-19 Vaccine and Social Media in the U.S.: Exploring Emotions and Discussions on Twitter', *Vaccines*, vol. 9, no. 10, p. 1059, 2021, doi: 10.3390/vaccines9101059.

[121] S. Lunn, J. Zhu, and M. Ross, 'Utilizing Web Scraping and Natural Language Processing to Better Inform Pedagogical Practice', *Proc. - Front. Educ. Conf. FIE*, vol. 2020-Octob, 2020, doi: 10.1109/FIE44824.2020.9274270.

[122] R. Proscia, 'Sentiment Analysis on Twitter', no. February, 2016.

[123] R. Vording, 'Harvesting unstructured data in heterogenous business environments ; exploring modern web scraping technologies', pp. 1–9.

[124] B. Naimi and A. Voinov, 'StellaR: A software to translate Stella models into R open-source environment', *Environ. Model. Softw.*, vol. 38, pp. 117–118, 2012, doi: 10.1016/j.envsoft.2012.05.012.

[125] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, 'A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis', *Expert Syst. Appl.*, vol. 110, pp. 298–310, 2018, doi: 10.1016/j.eswa.2018.06.022.

[126] G. Nelson, 'An Introduction to Corpus Linguistics', *J. English Linguist.*, vol. 28, no. 2, pp. 193–196, 2000, doi: 10.1177/00754240022004965.

[127] 'Identify and Remove Duplicate Data in R - Datanovia'. https://www.datanovia.com/en/lessons/identify-and-remove-duplicate-data-in-r/ (accessed Feb. 05, 2022).

[128] 'What's the Difference Between Emoji and Emoticons? | Britannica'. https://www.britannica.com/story/whats-the-difference-between-emoji-and-emoticons (accessed Feb. 06, 2022).

[129] 'delete.stop.words function - RDocumentation'. https://www.rdocumentation.org/packages/stylo/versions/0.7.4/topics/delete.stop.words (accessed Feb. 07, 2022).

[130] 'Stemming and lemmatization'. https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html (accessed Feb. 07, 2022).

[131] 'What Is Lemmatization? - Twinword'. https://www.twinword.com/blog/what-is-lemmatization/ (accessed Feb. 07, 2022).

[132] E. Nguyen, 'Text Mining and Network Analysis of Digital Libraries in R', *Data Min. Appl. with R*, pp. 95–115, Jan. 2014, doi: 10.1016/B978-0-12-411511-8.00004-9.

[133] '5.1 Tidying a document-term matrix | Notes for "Text Mining with R: A Tidy Approach"'. https://bookdown.org/Maxine/tidy-text-mining/tidying-a-document-term-matrix.html (accessed Feb. 07, 2022).

[134] S. Osailan, 'Topic modeling location-based social media applications', *Diss. Abstr. Int. Sect. B Sci. Eng.*, vol. 81, no. 12-B, 2020, [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2020-51430-205&site=ehost-live.

[135] İ. Kızıltan, 'An Evaluation of Istanbul Real Estate Posts Based on Text Analysis and Topic Modelling', 2018.

[136] M. Bhat, M. Kundroo, and T. A. Tarray, 'Deep LDA : A new way to topic model Deep LDA : A new way to topic model', no. July, 2019, doi: 10.1080/02522667.2019.1616911.

[137] K. Macmillan and J. D. Wilson, 'Topic Supervised Non-negative Matrix Factorization', pp. 1–20, 2017.

[138] E. Ozkan, 'Topic Modelling by Using Latent Dirichlet Allocation', 2015.

[139] I. Vayansky, S. A. P. Kumar, and Z. Li, 'An Evaluation of Geotagged Twitter Data during Hurricane Irma Using Sentiment Analysis and Topic Modeling for Disaster Resilience', *Int. Symp. Technol. Soc. Proc.*, vol. 2019-Novem, 2019, doi: 10.1109/ISTAS48451.2019.8937859.

[140] S. Kwon and A. Park, 'Understanding user responses to the COVID-19 pandemic on Twitter from a terror management theory perspective: Cultural differences among the US, UK and India', *Comput. Human Behav.*, vol. 128, no. May 2021, p. 107087, 2022, doi: 10.1016/j.chb.2021.107087.

[141] L. Zou and W. W. Song, 'LDA-TM: A two-step approach to Twitter topic data clustering', *Proc. 2016 IEEE Int. Conf. Cloud Comput. Big Data Anal. ICCCBDA 2016*, pp. 342–347, 2016, doi: 10.1109/ICCCBDA.2016.7529581.

[142] A. You, M. A. Y. Be, and I. In, 'Topic modeling Twitter data using Latent Dirichlet Allocation and Latent Semantic Analysis Topic Modeling Twitter Data Using Latent Dirichlet Allocation and Latent Semantic Analysis', vol. 020093, no. December 2019, 2021.

[143] D. M. Blei, A. Y. Ng, and M. I. Jordan, 'Latent Dirichlet Allocation David', *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[144] Z. A. Güven, R. T. Erdoğan, B. Diri, and T. Çakaloğlu, 'n-seviyeli Gizli Dirichlet Ayırımı ile Türkçe Tivit Duygularının Sınıflandırılması Classification of TurkishTweet Emotions by n-stage Latent Dirichlet Allocation', *Academia.Edu*, no. Asyu, 2012, [Online]. Available: https://www.academia.edu/download/59075094/PID528509720190429-80797-1c57sqz.pdf.

[145] 'Latent Dirichlet Allocation(LDA) Kullanarak Nasıl Topic Modeling Yapılır? | by Büşra Gökmen | Medium'. https://medium.com/@busragokmen67/latent-

dirichlet-allocation-lda-kullanarak-nasıl-topic-modeling-yapılır-75fe8dddcdd2 (accessed Mar. 12, 2022).

[146] ‘Gizli Dirichlet Tahsisi: pyLDAvis ile sezgi, matematik, uygulama ve görselleştirme | tarafından Ioana | Veri Bilimine Doğru’. https://towardsdatascience.com/latent-dirichlet-allocation-intuition-math-implementation-and-visualisation-63ccb616e094#7b8f (accessed Mar. 12, 2022).

[147] K. Para, ‘Investigation of the Causality Between Twitter Comments and Altcoin Cyptocurrency Market by Sentiment Analysis During the COVID-19 pandemic : Ripple Example’, no. September, pp. 362–381, 2021.

[148] O. Kaynar, M. Yildiz, Y. Görmez, and A. Albayrak, ‘Makine Öğrenmesi Yöntemleri ile Duygu Analizi’, no. September, pp. 234–241, 2016.

[149] D. Zimbra, A. Abbasi, D. Zeng, and H. Chen, ‘The State-of-the-Art in Twitter Sentiment Analysis’, *ACM Trans. Manag. Inf. Syst.*, vol. 9, no. 2, pp. 1–29, 2018, doi: 10.1145/3185045.

[150] A. Sarlan, C. Nadam, and S. Basri, ‘Twitter sentiment analysis’, *Conf. Proc. - 6th Int. Conf. Inf. Technol. Multimed. UNITEN Cultiv. Creat. Enabling Technol. Through Internet Things, ICIMU 2014*, pp. 212–216, 2015, doi: 10.1109/ICIMU.2014.7066632.

APPENDIX A

Appendix A1. The Code for Preprocessing and Topic Modelling Phase

```
"Installing related packages and libraries"

install.packages("tm")

library(tm)

library(textclean)

library(tokenizers)

install.packages("textstem")

library(textstem)

library(dplyr)


"for topic modelling"

library(tidyverse)

library(tidytext)

library(topicmodels)

library(quanteda)

library(wordcloud)

library(tidyr)

library(ggplot2)

install.packages('devtools')

devtools::install_github("lchiffon/wordcloud2")

library(wordcloud2)

install.packages("reshape2")


"Importing the dataset"

vaccination_data <- read.csv(file.choose(), header = T)
```

```
View(vaccination_data)


"Putting the data into a corpus for text processing"

tweet_corpus <- (VectorSource(vaccination_data$Tweet))

tweet_corpus <- Corpus(tweet_corpus)

inspect(tweet_corpus[1:131])


"~~~~~~~~PREPROCESSING PART~~~~~~~~"


"Removing duplications"

removeDup <- function(x) unique(x)

tweet_corpus <- tm_map(tweet_corpus, removeDup)

inspect(tweet_corpus[1530:1540])


"Converting characters to lower case"

tweet_corpus <- tm_map(tweet_corpus, content_transformer(tolower))

inspect(tweet_corpus[1:131])


"Removing URLs"

removeURL <- function(x){

  gsub("http[^[:space:]]*", "", x)

}

tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeURL))

inspect(tweet_corpus[1:131])


"Removing username"
```

```
removeUsername <- function(x){
  gsub("@[^[:space:]]*", "", x)
}
tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeUsername))
inspect(tweet_corpus[1:131])

"Removing hashtag"
removeHashtag <- function(x){
  gsub("#\\S+", "", x)
}
tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeHashtag))
inspect(tweet_corpus[1:131])

"Removing carriage"
removeCarriage <- function(x){
  gsub("[\r\n]", "", x)
}
tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeCarriage))
inspect(tweet_corpus[1:131])

"Removing emoticon"
removeEmoticon <- function(x){
  gsub("[^\x01-\x7F]", "", x)
}
tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeEmoticon))
inspect(tweet_corpus[1:131])
```

```
"Removing numbers"

tweet_corpus <- tm_map(tweet_corpus, removeNumbers)

inspect(tweet_corpus[1:131])


"Removing punctuation"

tweet_corpus <- tm_map(tweet_corpus, removePunctuation)

inspect(tweet_corpus[1:131])


"Removing anything except the English language and space"

removeNumPunct <- function(x) gsub("[^[:alpha:][:space:]]*", "", x)

tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeNumPunct))

inspect(tweet_corpus[1:131])


"Removing other invoices"

removeInvoice <- function(x){

  gsub("inv/[0-9]+/+[xvi]+/[xvi]+/[0-9]+", "", x, ignore.case = T)

}

tweet_corpus <- tm_map(tweet_corpus, content_transformer(removeInvoice))

inspect(tweet_corpus[1:131])


"Removing stopwords"

tweet_corpus <- tm_map(tweet_corpus, removeWords, stopwords("english"))

inspect(tweet_corpus[1:131])


"Lemmatization"

tweet_corpus<-tm_map(tweet_corpus, textstem::lemmatize_strings)
```

```
inspect(tweet_corpus[1:131])


"Eliminating words that were used as a keyword in Twitter search"

tweet_corpus <- tm_map(tweet_corpus, removeWords,c( "vaccine", "vaccination",
"biontech","sinovac", "vaccinal", "immunization",
                                   "covid", "corona", "coronavirus", "pandemic",
"epidemic", "children",
                                   "child", "kid", "boy", "daughter", "youngster", "junior",
"tiddler", "son", "parent",

"vaccinate","get","just","today","now","good","today","like","can","day","people","see"
,
                                   "take",
"still","give","make","know","tell","say","year","one","will","putin","russian","russia","
vladimir","putins"))
inspect(tweet_corpus[1:131])


"Eliminating extra whitespace"
tweet_corpus <- tm_map(tweet_corpus, stripWhitespace)
inspect(tweet_corpus[1:131])


"Extracting preprocessed data into csv file for further usage"
dataframe <- data.frame(text=sapply(tweet_corpus, identity),
                   stringsAsFactors=F)
df <- data.frame(text = get("content", tweet_corpus))
head(df)
write.csv(df, 'my.csv')
```

```
"Creating document-term matrix"

dtm <- DocumentTermMatrix(tweet_corpus)


"~~~~~~~~TOPIC MODELLING PART~~~~~~~~~~"


"Collapsing matrix by summing over columns"

dtm <- removeSparseTerms(dtm, sparse = 0.95)

frequency <- colSums(as.matrix(dtm))

"Length should be total number of terms"

length(frequency)

"Creating sort order (descending)"

ord <- order(frequency, decreasing = TRUE)

"Listing all terms in decreasing order of freq and write to disk"

frequency[ord]

write.csv(frequency[ord], "word_freq.csv")

"Removing the 0 rows"

raw.sum=apply(dtm,1,FUN=sum)

dtm=dtm[raw.sum!=0,]


"Creating model with 4 topics"

k=4

seed=1234

lda_fit <- LDA(dtm, k=k, control=list(seed=seed))

lda_fit@alpha

topics(lda_fit, k)

terms(lda_fit, 5)
```

```
"Tagging the docs to topics and topics to terms"
lda_fit.topics <- as.matrix(topics(lda_fit))
write.csv(lda_fit.topics,file=paste("docstotopics", k, "DocsToTopics.csv"))

lda_fit.terms <- as.matrix(terms(lda_fit,4))
write.csv(lda_fit.terms, file=paste("topicstoterms", k, "DocsToTerms.csv"))
lda_fit.terms[1:4,]

"Plotting a frequency graph based on topics' number"
top_terms_by_topic_LDA <- function(input_text,
                                   plot = T,
                                   number_of_topics = 4)
{

  topics <- tidy(lda_fit, matrix = "beta")

  top_terms <- topics  %>%
    group_by(topic) %>%
    top_n(5, beta) %>%
    ungroup() %>%
    arrange(topic, -beta)

  if(plot == T){
    # plot the top 5 terms for each topic in order
    top_terms %>% # take the top terms
```

```
    mutate(term = reorder(term, beta)) %>% # sort terms by beta value
    ggplot(aes(term, beta, fill = factor(topic))) + # plot beta by theme
    geom_col(show.legend = FALSE) + # as a bar plot
    facet_wrap(~ topic, scales = "free") + # which each topic in a seperate plot
    labs(x = NULL, y = "Beta") + # no x label, change y label
    coord_flip() # turn bars sideways
  }else{
   # if the user does not request a plot
   # return a list of sorted terms instead
   return(top_terms)
  }
}

top_terms_by_topic_LDA(tweet_corpus, number_of_topics = 4)

"Converting corpus for further manipulation"
tdm <- TermDocumentMatrix(tweet_corpus)
tdm <- removeSparseTerms(tdm, sparse = 0.95)
m <- as.matrix(tdm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)

"Plotting the most frequent words"
barplot(d[1:10,]$freq, las = 2, names.arg = d[1:10,]$word,
        col ="lightblue", main ="Most frequent words",
```

```
        ylab = "Word frequencies")


"Generating word cloud"
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 5,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
findFreqTerms(tdm, lowfreq = 4)


"Clustering associated words"
tdm2 <- removeSparseTerms(tdm, sparse = 0.95)
m2 <- as.matrix(tdm2)
distMatrix <- dist(scale(m2))
fit <- hclust(distMatrix, method = "ward.D2")
plot(fit)
rect.hclust(fit, k=4)
```

Appendix A2. The Code for Sentiment Analysis Phase

```
"Installing packages"
install.packages("tm")  # for text mining
install.packages("SnowballC") # for text stemming
install.packages("wordcloud") # word-cloud generator
install.packages("RColorBrewer") # color palettes
install.packages("syuzhet") # for sentiment analysis
install.packages("ggplot2") # for plotting graphs
install.packages("RCurl", repos = "http://cran.us.r-project.org")
install.packages("httr", repos = "http://cran.us.r-project.org")
```

```r
install.packages("syuzhet", repos = "http://cran.us.r-project.org")
install.packages("plotly")

"Loading libraries"
library(NLP)
library("tm")
library("SnowballC")
library("wordcloud")
library(wordcloud2)
library("RColorBrewer")
library("syuzhet")
library(stringr)
library("ggplot2")
library(reshape2)
library(janeaustenr)
library(dplyr)
library(stringr)
library(tidytext)
library(sentimentr)
library(plyr)
library(RCurl)
library(httr)
library(topicmodels)
library(tidyr)

"Importing the dataset"
```

```
vaccination_data_t1 <- read.csv(file.choose(), header = T)
View(vaccination_data_t1)

"Counting emotions from dataset with the help of relatd library"
d <- get_nrc_sentiment(vaccination_data_t1$Tweet)
head (d,10)

"Transposing"
td<-data.frame(t(d))
"Computing column sums across rows for each level of a grouping variable"
td_new <- data.frame(rowSums(td[1:5802]))
"Transformation and cleaning"
names(td_new)[1] <- "count"
td_new <- cbind("sentiment" = rownames(td_new), td_new)
rownames(td_new) <- NULL
td_new2<-td_new[1:8,]
"Plotting One - count of words associated with each sentiment"
quickplot(sentiment, data=td_new2, weight=count, geom="bar", fill=sentiment,
     ylab="count")+ggtitle("Sentiment scores of tweets")

"Plotting two - count of words associated with each sentiment, expressed as a percentage"
barplot(
  sort(colSums(prop.table(d[, 1:8]))),
  horiz = TRUE,
  cex.names = 0.7,
  las = 1,
```

```
  main = "Emotions in Text", xlab="Percentage"
)

"Calculating final score by using total positive and total negative"
d$score<-d$positive-d$negative
d[1:10,]

write.csv(x=d,file="C:/Users/elifd/Desktop/2021-2022/YL/Tez/Kod ve deneme belgeler/skor.csv")

"Calculating tweet scores in terms of different emotions"
tweetscore <- colSums(d[,])
print(tweetscore)

"Emotion classification & positive and negative sentiments"

nrc_average <- apply(d,2,mean)
nrc_average
sentisum <- colSums(d)
sentisum
Lb <- paste(names(sentisum), ",", sentisum)
pie(sentisum[1:10],col=brewer.pal(8,'Dark2'), labels=Lb,
    main="Emotions and Sentiment nrc Scores", cex=0.8, cex.main=2)

"YENİ"
emotions <- get_nrc_sentiment(vaccination_data_t1$Tweet)
```

```
emo_bar = colSums(emotions)

emo_sum = data.frame(count=emo_bar, emotion=names(emo_bar))

emo_sum$emotion = factor(emo_sum$emotion,
levels=emo_sum$emotion[order(emo_sum$count, decreasing = TRUE)])

# Visualize the emotions from NRC sentiments

library(plotly)

p <- plot_ly(emo_sum, x=~emotion, y=~count, type="bar", color=~emotion) %>%

  layout(xaxis=list(title=""), showlegend=FALSE,

         title="Emotion Type for vaccination")

p

wordcloud_tweet = c(

  paste(vaccination_data_t1$Tweet[emotions$anger > 0], collapse=" "),

  paste(vaccination_data_t1$Tweet[emotions$anticipation > 0], collapse=" "),

  paste(vaccination_data_t1$Tweet[emotions$disgust > 0], collapse=" "),

  paste(vaccination_data_t1$Tweet[emotions$fear > 0], collapse=" "),

  paste(vaccination_data_t1$Tweet[emotions$joy > 0], collapse=" "),

  paste(vaccination_data_t1$Tweet[emotions$sadness > 0], collapse=" "),

  paste(vaccination_data_t1$Tweet[emotions$surprise > 0], collapse=" "),

  paste(vaccination_data_t1$Tweet[emotions$trust > 0], collapse=" ")

)


corpus = Corpus(VectorSource(wordcloud_tweet))

tdm = TermDocumentMatrix(corpus)


# convert as matrix

tdm = as.matrix(tdm)
```

```
td_new <- cbind("sentiment" = rownames(tdm), tdm)

colnames(tdm) = c('anger', 'anticipation', 'disgust', 'fear', 'joy', 'sadness', 'surprise', 'trust')

comparison.cloud(tdm, random.order=FALSE,
        colors = c("#00B2FF", "red", "#FF0099", "#6600CC", "green", "orange", "blue", "brown"),
        title.size=1, max.words=250, scale=c(2, 0.4),rot.per=0.4)
```

CURRICULUM VITAE

Name Surname : Elif Delice

EDUCATION:

B.Sc. : 2020, Istanbul Medipol University, School of Engineering and Natural Sciences, Industrial Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2015-2020, Honorary Student, Istanbul Medipol University

OTHER PUBLICATIONS, PRESENTATIONS, AND PATENTS:

- E. Delice , B. Turan , A. Himtaş , H. Tozan and M. Karadayı , "Nontraditional Machining Process In Healthcare Applications", Journal of Health Systems and Policies, vol. 3, no. 1, pp. 20-41, Apr. 2021

PUBLIC PERCEPTION TOWARDS CHILDREN'S COVID-19 VACCINATION WITH NATURAL LANGUAGE PROCESSING

ORIGINALITY REPORT

10%	8%	3%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	acikerisim.medipol.edu.tr Internet Source	1%
2	devgin.com Internet Source	<1%
3	Submitted to Harrisburg University of Science and Technology Student Paper	<1%
4	www.msbiinfo.com Internet Source	<1%
5	Submitted to University of St Andrews Student Paper	<1%
6	towardsdatascience.com Internet Source	<1%
7	www.medicine20.com Internet Source	<1%
8	Submitted to University of Maryland, University College Student Paper	<1%