INTERNATIONAL
HELLENIC
UNIVERSITY

Social Media Sentiment Analysis Related to COVID-19 Vaccines: Case studies in English and Greek language

Kapoteli Evridiki
SID: 3308200012

SCHOOL OF SCIENCE & TECHNOLOGY
A thesis submitted for the degree of
Master of Science (MSc) in Data Science

JANUARY 2022
THESSALONIKI – GREECE

INTERNATIONAL HELLENIC UNIVERSITY

Social Media Sentiment Analysis Related to COVID-19 Vaccines: Case studies in English and Greek language

Evridiki Kapoteli

SID: 3308200012

Supervisor: Prof. Christos Tjortjis

Supervising Committee Members: Dr K. Tzafilkou

Dr D. Karapiperis

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in Data Science

JANUARY 2022

THESSALONIKI – GREECE

Abstract

SARS-CoV-2 and its mutations are rapidly spreading around the world threatening human population with millions of infections and deaths. Vaccines are the only available weapon at hand to mitigate the spread. As a result, the development of efficient systems to understand and supervise the information dissemination as well as the sentiment evolution toward vaccines is critical. The goal of this research was to build and apply a supervised machine learning approach to monitor the dynamics of public opinion on COVID-19 vaccines using Twitter data. 1,394,535 and 61,077 tweets in English and Greek about COVID-19 vaccines, respectively, were collected, classified based on sentiment polarity, and analyzed over time to gain insights into sentiment trends. The findings reveal that overall negative, neutral, and positive sentiments were at 36.5%, 39.9% and 23.6% in the English language dataset, respectively, whereas overall negative and non-negative sentiments were at 60.1% and 39.9% in the Greek language dataset. Policy makers and health experts should take into consideration social media sentiment analysis alongside other ways of evaluating public sentiment. Social media users are actively seeking and sharing information about all pandemic-related topics, allowing governments to use social media not only to better inform the public with accurate and reliable news, but also alleviate disease-specific concerns, minimize the distribution of fake news, and develop effective crisis management strategies.

Keywords: Sentiment Analysis, COVID-19, social media, vaccine, natural language processing, Twitter, Supervised learning, classification, text mining, coronavirus, machine learning

Evridiki Kapoteli

01/07/2022

Acknowledgments

I would like to start by thanking my supervisor Prof. Christos Tjortjis for his invaluable guidance and encouragement throughout the duration of my dissertation. He has always been supportive, providing me great advise, and resources, as well as constructive comments and suggestions. Furthermore, I would like to express my gratitude to the International Hellenic University without which the selection and completion of this dissertation would not be possible. Finally, I am extremely grateful to my family and friends for their patience and support all this time.

Contents

ABSTRACT **3**

ACKNOWLEDGMENTS **4**

CONTENTS **5**

1 INTRODUCTION **7**

2 LITERATURE REVIEW **10**

2.1 Sentiment Analysis 10

2.2 Sentiment Detection Approaches 10

2.3 Sentiment Analysis for Multilingual Documents 11

2.4 Sentiment Analysis for Modern Greek Documents 17

3 RESEARCH DESIGN **21**

3.1 Methodology 21

3.2 COVID-19 Vaccine Dataset 23

3.2.1 Dataset Collection *23*

3.2.2 Dataset Annotation *25*

3.2.3 Dataset Pre-processing *27*

3.3 Text Representation 28

3.4 Learning Algorithms 29

4 EXPERIMENTAL RESULTS **31**

4.1 English Language Data 33

4.2 Greek Language Data 33

4.3 English Language Dataset Trend Analysis 35

4.4 Greek Language Dataset Trend Analysis 40

5 CONCLUSIONS **44**

5.1 Discussion 44

5.2 Limitations And Future Work 45

BIBLIOGRAPHY 47

1. Introduction

In mid-December 2019, the first cases of the coronavirus outbreak were emerging with humanity being completely unaware, until March 11, 2020, when WHO declared a global pandemic [1] and concern escalated around the globe. Since then, 220 countries have been affected, with nearly 277 million confirmed cases and over 5,5 million deaths reported [2]. Due to the emergency, national governments were forced to consider a series of stringent measures aimed at preventing SARS-CoV-2 from spreading across communities. The scientific community, on the other hand, took swift action and made an unprecedented effort towards developing safe and effective vaccines to combat the threat.

People from their side, being asked and encouraged to stay at home, turned towards social media to stay informed and to communicate their thoughts. Especially in early 2020, when humanity was dealing with widespread lockdowns, social distancing and closure of businesses and schools, social media was one of the most reliable sources of the latest coronavirus information. Keeping up to date with the most recent news and information throughout this period was of outmost importance to social media users, often at the cost of their mental health, resulting in anxiety and depression [3]. Apart from serving as a medium of information, flooded in virus- and vaccine-related news, social media has evolved into an important tool for interacting with friends and family, as well as for expressing ideas and concerns with people worldwide [4].

All of the aforementioned factors have contributed to the extensive use of popular social platforms, demonstrating that social media may be harnessed to foresee events related to the global spread of the disease in real-time [5]. Among some of the leading social networks globally, Twitter counts a large number of active users as it constitutes a great place to broadcast opinions and thoughts on topics of public interest through short-length posts called tweets. Twitter is regarded by the research community as one of the most valuable sources of user-generated datasets. In previous years, Twitter data has been used to infer insights on a plethora of topics, including vaccination.

Wearing masks and maintaining social distance have been the only short-term preventive measures towards confronting the pandemic we had all this time. However, now that a number of vaccines have been developed, the long-term containment of the pandemic depends solely on their uptake. As of December 2021, there are 276 vaccines in development that fall into 9 different product categories, while 107 of them are in clinical testing and 24 vaccines are already in use [6]. However, the novelty of the disease and worries regarding efficacy, safety, side effects, and vaccine development speed, as well as poor or insufficient communication, all contribute to the population's unwillingness to receiving the COVID-19 vaccine. Vaccine hesitancy has always been a source of concern for humanity, and the World Health Organization has placed it among the major threats to global health [7]. The COVID-19 emergency is no exception.

Due to the rapid spread of coronavirus and its variants, there is an urgent need for developing efficient systems in order to understand and monitor the flow of information and the evolution of sentiment. This research is motivated by the extensive use of Twitter by people all around the world during the COVID-19 outbreak and intends to analyze Twitter data for monitoring public opinion regarding COVID-19 vaccines during that time. We consider the 7-month period between May 19, 2021 and November 19, 2021 to collect Twitter messages written in English and Greek. During the study period, we extracted 1,394,535 English tweets and 61,077 Greek tweets, which were reduced to 1,257,944 and 50,796 respectively after cleaning. The sentiment analysis performance of various machine learning algorithms was evaluated using an annotated dataset for each language, and the best performing classifier was chosen and applied to the entire dataset. The sentiment analysis task was addressed as a binary task of classifying sentiment into negative and non-negative classes for Greek language text, and as a 3-way classification problem for English language text, with negative, neutral, and positive classes.

Facebook, Twitter, and Instagram are three of the world's most popular social networking services. In our research, we examine how Twitter may be used for sentiment analysis during the coronavirus crisis. The following are some of the reasons we preferred Twitter as our main data source.

First and foremost, Twitter data is rich in information and can be readily and freely accessible by anyone with the appropriate usage rights. This data is retrieved via the provided Application Program Interface, which is well-documented and simple to use. Additionally, because Twitter has a massive volume of text posts that is ever-growing, the retrieved collection can be as large as desired. Throughout the pandemic, the platform served as a channel for exchanging personal thoughts, feelings, and information about a variety of topics, making it a valuable source of public opinion. More interestingly, Twitter is a medium that is used not only by ordinary users but also by celebrities, politicians, health professionals, and reputable health agencies, like WHO. Therefore, tweets written by users from various social and interest groups can be collected. Additionally, Twitter users might originate from many countries around the world. Although the majority of the users are from the United States, Japan, and India [8], data is available in a variety of languages. Finally, and most crucially, Twitter has actively worked to prevent misinformation from spreading by eliminating tweets whose content opposes to global or local authority guidelines [9].

Because there are limited linguistic resources for Greek NLP tasks and little research on COVID-19 vaccines in general, this work provides a promising solution, as illustrated by the experimental results below. Following that, we highlight our research's key contributions:

- A collection and annotation of two COVID-19 vaccination datasets, one in English and one in Modern Greek. To our knowledge, this is the first Modern Greek Twitter dataset about COVID-19 vaccines.

- A comparative evaluation of how sentiment analysis and opinion mining methods work on Modern Greek, an under-resourced language for NLP tasks where sentiment analysis is rare, and English, the language on which most research in this area is conducted.
- A comparison of state-of-the-art sentiment classification approaches including classical machine learning models as well as pre-trained language models. The determination of the best performing model for COVID-19 vaccines sentiment analysis.
- A comparative analysis of social media opinions and sentiment towards the COVID-19 vaccination among Greek individuals and the global community.

The following is an outline of the remaining research. Section 2 includes an overview of the literature on natural language processing with a focus on sentiment analysis, followed by a discussion of recent studies employing Twitter to perform sentiment analysis on COVID-19. Section 3 describes in detail the problem and the proposed research design, addressing data collection, annotation, pre-processing and representation processes, as well as the learning algorithms and their experimental results. Following that, Section 4 focuses on analyzing the experimental results by detailing the performance of the machine learning algorithms used as well as examining the sentiment trend as expressed on social media. The final sections contain the dissertation's conclusions, limitations, and future directions, as well as a list of references.

2. Literature Review

This section provides a concise survey of sentiment analysis, text analytics, Twitter, and NLP literature, in order to highlight the existing research methodologies. Following that, we discuss a collection of studies that leveraged Twitter data to analyze public opinion on emergent topics, including vaccines, focused on the Greek and English language during the COVID-19 pandemic.

2.1 Sentiment Analysis

Opinion mining and sentiment analysis are two terms that are sometimes used interchangeably to refer to the same research area, however the former focuses on polarity detection and the latter on emotion recognition. Both tasks require the use of techniques like data mining and natural language processing to identify, collect, and distill information and opinions from the massive textual content available online [10]. The advantage of sentiment analysis is that it enables any organization to track multiple social media platforms in real time and react appropriately. Thus, there exist sentiment analysis applications in nearly every field, including products, healthcare, as well as finance and political elections.

Sentiment analysis can be applied to text at three different levels: document, sentence, and aspect levels. In document-level sentiment analysis, we assume that each document represents the author's opinion on one main entity. Sentence-level sentiment analysis examines individual sentences in a document to identify if they convey a positive, negative, or neutral opinion. This level of analysis is extremely useful when multiple perspectives are included in a single document, even though they all relate to the same entities. However, it is possible that we have a document in which multi-aspect entities are discussed and different opinions are given on each aspect. Then, aspect-based sentiment analysis is applied, which is focused on identifying the sentiment expression for each of the existing aspects [12].

2.2 Sentiment Detection Approaches

There are three basic approaches that have been employed in the Twitter Sentiment Analysis literature: machine learning methods, lexicon-based methods, and hybrid methods. In order to construct a model that can classify tweets conveying opinions or sentiments, the machine learning technique begins by dividing the textual resources into training and testing sets. The classifier learns from training data to identify features based on which then it classifies previously unseen data into a collection of predefined sentiment categories [13]. Limitations of this approach include the reliance on the number of training data, meaning that obtaining a good performance requires a large amount of labeled data. Furthermore, domain-dependency is a limitation indicating that applying a classifier to a domain other than the one in which it was trained may reduce its performance. On the other hand, the lexicon-based technique finds the overall sentiment score of a tweet by leveraging a list of terms from the

desired language labeled by polarity or polarity score. This unsupervised approach has the advantage of not requiring training data, but it suffers when the given text contains informal language, such as that seen on social media, because the lexicon lacks context-specific phrases [14]. Hybrid is a performance-improving technique that blends machine learning with lexicon-based methods to overcome the limits of each technology when employed individually [11].

The difficulty in assessing opinions and sentiments from human language arises from the need for a deep understanding of all language rules. It is important to remember that sentiment analysis is an NLP problem, and as such, it forces the research community to deal with unresolved issues like coreference resolution, negation handling, and word-sense disambiguation [12] [10]. Mining tweets' sentiments and opinions is significantly more challenging because of the special features of Twitter. Twitter Sentiment Analysis addresses the problem of identifying sentiments expressed in Twitter posts. The length constraint and unstructured and informal nature of the medium results in the use of incorrect English like emphatic uppercasing or lengthening and the use of slang. Other difficulties that researchers attempting to build effective Twitter Sentiment Analysis models face include data sparsity, which can impact the overall performance of the sentiment analysis model, as well as stop words, which are frequently used words with low discrimination power. Finally, equally critical to determining a tweet's sentiment polarity is the existence of negation words. Negations must be detected and handled appropriately, since they can cause the polarity of a message to be reversed [15].

In the discussed framework, useful information is extracted from tweets using text mining techniques along with methods from data mining, machine learning, statistics, and Natural Language Processing. Text mining is a technique for automatically extracting information from unstructured natural language corpora. The ambiguity of natural language acts as an obstacle to text mining and is due to the widespread usage of idioms, slang expressions, grammatical variations, and abbreviated words [16].

In the subsections that follow, a few recent studies for each category of techniques for each considered language are recalled, with an emphasis on Twitter messages' analysis.

2.3 Sentiment Analysis for Multilingual Documents

Sentiment Analysis on English-language text is a growing study area that has received a lot of interest. In their study [17], Pak & Paroubek collected Twitter data and used it to build a sentiment analysis model that can discern between positive, negative, and neutral sentiments expressed in a document. The corpus included 300,000 English-written tweets that fit into one of three categories. To collect documents reflecting positive and negative sentiments, the writers queried Twitter for happy emoticons and sad emoticons, respectively. The training set for the neutral class was acquired by querying Twitter accounts of popular newspapers. To build the sentiment model, different classifiers were examined, including SVM and CRF,

but Naïve Bayes outperformed them all. Thus, two Bayes classifiers were trained, each one using different features extracted from the collected corpus. Aiming to improve the classification accuracy, the authors considered of discarding n-grams that do not clearly reflect any sentiment or sentence objectivity and proposed two strategies to do so. In the final step, the classifier is tested on a manually annotated corpus of real tweets.

Twitter data has been extensively used for research in the past few years, the more so because of the emergency imposed by COVID-19. Mansoor et al. [18] collected three Twitter datasets, one related to coronavirus to examine sentiment evolution over time across different countries, and the other two related to work from home and online learning to evaluate the influence of the pandemic on daily life aspects. Exploratory data analysis was performed not only on the three aforementioned datasets but also on an additional dataset that includes the number of confirmed cases and deaths on a daily basis, in order to compare the sentiment change with the case number evolution from the start of the pandemic through June 2020. The tweets' sentiment content was identified using VADER, a rule-based sentiment analysis tool, and the sentiment classification was performed using deep learning models like Long Short-Term Memory and Artificial Neural Networks. The former achieved an accuracy of 84.5% on the coronavirus Twitter dataset tagged with VADER, whereas the latter achieved an accuracy of 76%.

Emotion analysis was performed on the same dataset, and it was discovered that the emotion of fear was stronger than that of trust throughout the study period, with the highest fear scores coming from Thailand, Vietnam, and Poland. On the contrary, the highest levels of trust were observed in Oman, Syria, and Kazakhstan. Finally, the sentiment analysis task revealed that Bangladesh, Pakistan, Mali, and South Africa were among the countries where positive sentiment was the dominant emotion, while Australia, India, Canada, the United States, Turkey, the United Kingdom, and Brazil had a greater proportion of negative sentiments.

In their study, Hung et al. [19] used English-language tweets stemming solely from the United States during a 1-month period and performed sentiment analysis and social network analysis to examine COVID-19 discussion themes and sentiments. Sentiment analysis using VADER found that positive tweets dominated, comprising 48.2% of the corpus, followed by negative (31.1%) and neutral (20.7%). Interestingly, the authors reported that "negative sentiment toward COVID-19 was more prevalent in sparsely populated states with lower infection rates." Similarly, Wang et al. [11] leveraged COVID-19 tweets to monitor and compare the public's sentiment and discussed topics about COVID-19 in two different states of America. The tweets' sentiment score was obtained by conducting sentiment analysis using VADER, while the most discussed topics were found by employing topic modeling. The findings indicate that there were changes in sentiment and that "protective measures" was the most prevalent topic in online discussions on social media.

Aiming to understand and examine public sentiment about social distancing during the first outbreak of the pandemic, Shofiya & Abidi [20] collected 629 Canada originated tweets

containing social distancing keywords and performed hybrid sentiment analysis. More specifically, the sentiment polarity of each tweet was extracted utilizing the SentiStrength tool and used in the model's training. The applied SVM model, being 87% accurate, revealed that 40% of the collected tweets conveyed neutral sentiments about social distancing, 35% showed negative sentiments, and only the remaining 25% showed positive sentiments.

Boon-Itt & Skunkan [21] also utilized sentiment analysis, topic modeling, and keyword frequency to better understand public perception of COVID-19 and to uncover topics of online discourse. For that purpose, the authors leveraged Twitter data posted in English during the study period. The majority of the tweets were found to express negative sentiments, indicating in that way the negative perception Twitter users had regarding the coronavirus pandemic. An in-depth analysis of the emotional content indicated that over half of the tweets expressed fear, trust, or anticipation. Moreover, it was found that emotions changed dynamically, and the researchers reported that events like the infection of more than 100,000 people with COVID-19 increased negative sentiment, whilst the availability of knowledge on prevention and protection increased the positive sentiment of the public.

Another study [22] attempted to track how Twitter users' behavior altered during the COVID-19 outbreak. The authors collected tweets from three different time periods using popular hashtags related to the coronavirus and performed sentiment and emotional analysis. The 'Negative' category had the largest percentage of tweets, while 'Sadness' was the most prevalent among other emotions, indicating that the majority of the population had negative feelings. The Twitter dataset was intergraded with WHO datasets about the pandemic's spread and mortality rates. The findings suggest that there were strong correlations between infection and death rates, as well as the emotional tendencies of Twitter users.

Twitter data for sentiment analysis has been used by Samuel et al. [23] to monitor public sentiment during the pandemic, based on coronavirus tweets. The authors examined the efficiency of the Naïve Bayes and Logistic Regression algorithms in classifying coronavirus tweets of different lengths as either positive or negative. Both algorithms were better at classifying short to medium-length tweets than longer ones, with the Naïve Bayes approach outperforming Logistic Regression. Further calculations showed that Naïve Bayes performed better when categorizing negative tweets, whereas logistic Regression performed better when the classification was balanced. The researchers also employed descriptive textual analytics to demonstrate the progression of fear, which was the most prevalent emotion in the Twitter corpus.

Furthermore, the Twitter flow of information during the COVID-19 outbreak was studied by Kaila et al. [24], who extracted tweets with the hashtag "#coronavirus" and conducted sentiment analysis and topic modeling. The Latent Dirichlet Allocation analysis proved successful in identifying relevant and valid topics related to the on-going situation. In addition, sentiment analysis revealed the presence of both positive and negative sentiments, with the latter dominating the corpus and trust coming in second.

Another study [25] used two distinct datasets to look at how people felt about COVID-19 on Twitter. The first was composed of the most retweeted tweets, while the second was comprised of a number of collected tweets. The analysis of the latter revealed that internet users tweeted the most positive and neutral messages, while the analysis of the former revealed that users retweeted the most tweets expressing negative or neutral opinions. Thus, despite the fact that the majority of tweets on COVID-19 were positive, people were more interested in retweeting the negative ones. These findings indicate how meaningless information on social media can spread as fast as the pandemic and have a harmful impact on society, especially during emergencies like COVID-19.

Vaccination has always been an emotionally charged topic for societies, even before the appearance of the coronavirus pandemic, and as a consequence, a substantial amount of research has been done on the subject. Yuan and Crooks [26] exploited sentiment analysis and machine learning techniques to investigate how anti- and pro-vaccination Twitter users interacted about the MMR vaccine. Furthermore, Raghupathi et al. [27] examined online sentiment, word usage, and opinions towards the measles vaccine, combining text mining and sentiment analysis techniques on social media. Du et al. [28] suggested a hierarchical machine learning-based methodology to analyze public sentiment on HPV vaccine-related tweets and assessed its performance on a large-scale dataset [29].

D'Andrea et al. [16] developed a system for dynamically inferring the public's stance towards immunization in Italy. Despite the fact that this work leverages Italian language tweets rather than English language tweets, we included it in the literature survey because it is worth mentioning. To fine tune the system, several experiments were conducted to determine the best performing combination of text representation and classification approach, using metrics like accuracy and F-measure. It was found that the combination employing the bag-of-words scheme along with a support vector machine model had the highest accuracy. Hence, based on a collection of keywords, vaccine-related tweets were retrieved from Twitter and preprocessed before being fed into the SVM model, which can predict the class of the tweet, namely in favor, not in favor, and neutral, being 64.84% accurate. The authors also reported the findings of a 10-month monitoring analysis of how the Italian citizens' stance on vaccination shifts in response to local spikes in the daily number of Twitter posts, which may be attributed to certain events relevant to the vaccine topic.

The creation of several life-saving vaccines has been the most significant advancement since the coronavirus appeared, and there have been many systematic reviews studying how people viewed their discovery. Beginning on March 1, 2020 and ending on February 28, 2021, Hu et al. [30] investigated public perception of COVID-19 vaccines in the United States. They categorized the study period into three distinct phases and examined how sentiments and emotions changed across time and space. Sentiment analysis using VADER, as well as emotion analysis and topic modeling methods, were applied on over 300,000 geo-tweets. The findings indicate that in the majority of the states, there was a rise in positive sentiment

accompanied by a drop in negative sentiment, and trust and anticipation were identified as the most notable feelings, along with fear, sadness, and anger feelings.

Lyu et al. [31] examined public discourse on social media around the COVID-19 vaccines to see what topics and sentiments were prevalent and how they changed during the study period. The required tweets were obtained from a COVID-19 Twitter dataset, and R software was used to perform topic modeling, sentiment, and emotion analysis. The latent Dirichlet allocation approach returned 16 topics, all of which belonged to one of five main themes, the most tweeted of which was "opinions on vaccination." When considering the overall sentiment, it was shown that it was becoming increasingly positive during the study period, peaking in November 2020. When emotions were examined, it was discovered that the most common emotion was trust, while anticipation and fear followed next. In fact, it was noted that as trust grew, fear decreased, and that, aside from these changes, the other emotions remained relatively stable during the research.

Cotfas et al. [32] compared classical machine learning and deep learning algorithms on a labeled dataset to determine which one captured the public perception of the new coronavirus vaccines the best. The authors gathered and studied a dataset containing tweets written in English and linked them to events discussed in the media over a period of time, starting with the initial vaccine announcement and ending with the beginning of the vaccination process in the UK. In order to train the models, a sample of the gathered dataset was chosen at random and was manually assigned into one of the following categories: in favor, against, or neutral. The performance of algorithms like Multinomial Nave Bayes, Random Forest, Support Vector Machine, Bidirectional Long Short-Term Memory, and Convolutional Neural Networks was examined using Accuracy, Precision, Recall, and F-score measures.

According to the research, classical machine learning classifiers outperformed deep learning classifiers, while the BERT language model, with an accuracy of 78.94%, was the classifier achieving the highest performance. Thus, the analysis is conducted using the BERT classifier and demonstrates that the predominant stance, either daily or entirely is neutral, whereas in favor tweets, outnumbered against tweets. Finally, the n-grams analysis identified a link between the tweets content and the media news, with the authors observing that "the occurrence of tweets follows the trend of the events."

Pristiyono et al. [33] performed Twitter sentiment analysis in their study to explore Indonesians' opinions on the COVID-19 vaccine in the second and third weeks of January 2021. A number of Indonesian tweets were collected and labeled before being used in the sentiment analysis model, which employed the Naïve Bayes classification algorithm. It was found that 56% of the tweets were negative, 39% were positive, and the remainder 1% were neutral. However, it was not reported how accurate the model's predictions were. In their work, Villavicencio et al. [34] performed sentiment analysis on COVID-19 vaccine-related tweets collected during the first month of vaccination in the Philippines. The authors manually labeled the gathered and preprocessed tweets (993 tweets) as positive, neutral, or

negative and used them to train a Naïve Bayes model. Following training and testing, the model's accuracy was equal to 81.77%.

Similarly, another study [35] employed an open-source dataset, comprised of COVID-19 vaccine tweets from all around the world, to determine the public stance about vaccination. The sentiment polarity of each tweet was extracted using TextBlob, and the authors chose to use only positive and negative tweets. For the sentiment analysis task, Multinomial Naïve Bayes (MNB), Support Vector Machine (SVM), and Logistic Regression (LR) classifiers were evaluated. The LR model performed the best, with an accuracy of 97.3%, followed by SVM and MNB, which had accuracies of 96.26% and 88%, respectively.

Public opinion and online discourse related to the COVID-19 vaccines were also studied in [36], where an artificial intelligence-based approach classified sentiment expressed in social media posts as positive, neutral, or negative and monitored its spatiotemporal evolution. The authors examined social media posts from the United Kingdom and the United States, extracted from Facebook and Twitter, for a study period of nine months. Using a rule-based ensemble, the weighted averaged result of Valence Aware Dictionary for Sentiment Reasoning and TextBlob was integrated with the findings supplied by the Bidirectional Encoder Representations from Transformers. The researchers then annotated a random portion of the posts by hand and fed them back into the model to validate and improve it. The findings indicate that posts containing positive sentiment were better identified by the lexicon-based methods, while negative and neutral content was identified with the highest accuracy by BERT. When the model was applied to the data, the average public perception about COVID-19 vaccinations was found to be mainly positive and consistent across Facebook and Twitter, in both countries. The use of word clouds and n-grams to examine spikes in the sentiment evolution graph can help uncover topics of debate and link them to positive or negative content in posts.

Sattar & Arifuzzaman [37] used various keywords to collect English written vaccine-related tweets as well as tweets about health and safety concerns following vaccination, aiming to examine public perception towards vaccination against COVID-19. The authors also tried to forecast the percentage of US citizens that will be fully vaccinated and will receive at least one dose of vaccine. An unsupervised lexicon-based method was chosen in which the performance of TextBlob and Vader sentiment analysis tools was compared. It was found that most of the tweets were neutral, and in the remaining part, positive sentiment outnumbered negative, despite reports of adverse reactions to some of the vaccines. Likewise, positive sentiment was higher than negative when it came to maintaining COVID-19 safety precautions after being vaccinated.

In the same direction, Yousefinaghani et al. [38], studied people's sentiments and opinions in relation to COVID-19 vaccines, using a Twitter dataset. The authors employed Valence Aware Dictionary and sEntiment Reasoner (VADER) to categorize each tweet as positive, negative, or neutral. The majority of the tweets belonged to the neutral category, while the

positive and negative categories contained 34% and 25% respectively. The study found that vaccine-related events had an impact on the trend of specified thoughts and sentiments, and users were objective and hesitant rather than interested in the vaccines, a behavior that differed by country.

Another study [39] used the Twitter API to extract tweets in English about the AstraZeneca/Oxford, Pfizer/BioNTech, and Moderna vaccines over a four-month period beginning December 1, 2020, in order to monitor the sentiment towards COVID-19 vaccines. The authors performed a lexicon-based sentiment analysis, employing the AFINN lexicon, and identified potential events and news that may have caused the sentiment to change. During the study period, the sentiment toward the Pfizer/ BioNTech and Moderna vaccines has been positive and stable, while the sentiment toward the AstraZeneca/ Oxford vaccine appeared to be decreasing.

2.4 Sentiment Analysis for Modern Greek Documents

Despite the fact that sentiment analysis of English-language text collections has evolved into a prominent research topic in recent years, as evidenced by the several works listed above, there has been relatively little work published on sentiment analysis of Greek language text collections. This could be due to the difficulty of adapting and applying existing NLP solutions to a language with complex morphological features like Modern Greek and many other European languages. Even at a smaller scale, there have been studies on the challenges that arise when constructing and applying a sentiment analysis model on Greek, some of which we discuss below.

Athanasiou & Maragoudakis [40], not only examined the application of machine learning in sentiment analysis tasks for under-resourced languages like Modern Greek, but also dealt with a class imbalance problem. User opinions posted in a Greek newspaper's online articles, covering a variety of topics, were collected while an amount of them were also annotated. However, the corpus was found to be imbalanced, and as stated by the authors, “a hybrid approach was followed that performs oversampling the minority class using the SMOTE (Synthetic Minority Oversampling TEchnique) algorithm and under-sampling the majority class using the Tomek links score.” Further, the suggested approach takes each Greek token's translation into account as an extra input feature in the training data’s feature set. During the development of the sentiment analysis model Gradient Boosting Machines (GBM) algorithm was used, which was found to effectively deal with imbalanced datasets of high dimension and outperform other state-of-the-art classifiers.

For the purpose of document level sentiment analysis, a dataset of tourism-related reviews written in Greek was collected, in order to be used for training and validation. The authors’ aim was to use the review text to predict the review’s score and classify it into one of five customer satisfaction categories. Each review was accompanied by a score ranging from 1 to 5, as determined by the reviewer, and was also preprocessed using a linguistics tool that labels each sub-sentence of the review with sentiment qualifiers depending on the encountered

terms. The authors then developed four distinct Deep Learning network designs and examined if raw text, text plus tag annotation information, or just the text's annotation was more helpful at assisting the model to categorize the sentiment expressed in a review. According to the findings, the creation and use of the tag annotation resulted in simpler but more accurate architectures and provided the same or higher accuracy when compared to traditional sentiment analysis [41].

Markopoulos et al. [42] compared the TF-IDF bag-of-words model with the term occurrence strategy to create a sentiment classifier for Greek language hotel reviews using a unigram language model. The researchers collected a dataset of 1,800 hotel reviews to train an SVM model that can discern between positive and negative sentiment polarity. The RapidMiner software was used for all experiments, and the results showed that the TF-IDF bag-of-words approach outperformed the term occurrence method.

Spatiotis et al. [43] in their study also presented the creation of a supervised sentiment analysis model to classify user-generated comments about e-lectures into one of the different classes, namely: very negative opinion, negative, neutral, positive, and very positive opinion. To this end, different classification algorithms were investigated and compared in terms of their accuracy, and it was found that for classes with a small number of training examples, the accuracy rate was not as satisfactory as for other classes.

Giatsoglou et al. [44] presented a framework for sentiment prediction of opinionated documents in different languages. A machine learning approach was followed, in which the vector representation of the documents was derived by combining a lexicon-based and a word embedding-based approach into hybrid vectors. Experiments were conducted on four datasets, two in English and two in Greek, comprised of online user reviews, with the goal of comparing the effectiveness of the proposed hybrid representations in detecting sentiment to employing the lexicon-based or word embedding-based approaches independently. Multiple hybrid vector generation strategies, as well as various coding schemes and classifiers, were tested and evaluated on the aforementioned datasets. Afterwards, the resultant representations were fed to the classification algorithm, with an SVM model proving to be the most accurate when a linear kernel was used. Finally, the results showed that hybrid vectors usually outperformed lexicon-based and word embedding-based methods without considerably increasing the model's needed running time.

Furthermore, Naïve Bayes, Random Forest, Support Vector Machines, Logistic Regression, and deep feed-forward neural networks were examined for sentiment classification. A binary-classification task, which included only positive and negative documents, and a multi-class classification task, which included examples from all three classes, were addressed. When considering the case of text representations, the research team first trained its own new model directly on a large social media corpus written in Greek and then trained an existing language model, GreekBERT, on the same data. The findings indicate that pre-trained language models performed better than traditional representation models and that they performed even better

when a proper language model was trained on a smaller yet domain and task relevant corpus. The selected combination of language model and classification algorithm can also have a significant impact on the overall performance.

Alexandridis et al. [45] employed text representation and text classification techniques and assessed the performance of different combinations of language models and classifiers in detecting the sentiment polarity of opinions expressed in Greek documents from social media. The training corpus consisted of text collected from various social media platforms that had been manually annotated by individuals. To map the retrieved text onto a vector space, the authors used three different pre-trained Greek language embeddings, namely BERT, FastText, and GPT-2.

Focusing also on Twitter Sentiment Analysis applied to Greek texts, Kalamatianos et al. [46] compared different approaches for extracting the sentiment rating of tweets and hashtags using a Greek Sentiment Lexicon. In the process, the authors developed and made public a collection of Greek tweets, as well as a manually annotated collection of tweets, which was used to evaluate the effectiveness of each approach. They also looked at how the intensity of "Happiness" and "Anger" for each hashtag changed over time in response to events.

More importantly, the work of Tsakalidis et al. [47] enabled public access to a manually annotated lexicon for the Greek language and the development of word embedding vectors and annotated datasets that can be used for a variety of tasks. The resources' performance was demonstrated through tasks including sentiment analysis, emotion and sarcasm detection. The authors achieved a higher F-score using their resources to train various algorithms than when using other existing methodologies for sentiment analysis.

Another study [48] used a combination of Greek lexicons and classification methods to determine how various political events influenced the emotions of Greek Twitter users in the days prior to the elections. A corpus of annotated tweets was preprocessed to create features, and each tweet was classified using a probabilistic classifier and a hashtag-based filter. This sentiment analysis model was then applied to a newly created dataset of tweets about the Greek elections in order to extract their sentiment.

Likewise, in the field of politics, Antonakaki et al. [49] used two distinct datasets taken from Twitter regarding two related political events: the Greek bailout referendum and the following legislative elections. The authors manually constructed a new political-domain lexicon for the Greek language and employed it in conjunction with SentiStrength's built-in lexicon and SocialSensor lexicon for the task of entity and sentiment detection. Following that, tasks like volume and sentiment analysis, as well as sarcasm correction and topic modeling were conducted to further their research.

[50] describes a supervised learning-based approach that classifies Greek language tweets as belonging in the negative, positive, or neutral class, depending on their sentiment. The authors extracted Greek tweets and manually labeled them in order to train and test the

proposed method, which was then compared to three other sentiment analysis algorithms established for English documents.

In the context of sentiment analysis using COVID-19 data, Kydros et al. [51] collected Twitter data, based on a set of keywords, to understand Greek citizens' feelings throughout four different time periods during the pandemic's first wave in Greece. The authors used NodeXL Pro to identify word pairs within tweets and to perform a lexicon-based sentiment analysis. Throughout all four study periods, the most discussed word pair was "new cases." For the sentiment analysis task, the lexicons provided by Tsakalidis et al. [47] were enriched with specific coronavirus-related words in order to be used in this study. The results indicate that sentiment fluctuated over time, fear dominated other emotions, and positive feelings declined while negative ones increased.

The only existing work in the Greek literature concerning COVID-19 vaccines is that of Kourlaba et al. [52], which aimed to examine how willing people were to get vaccinated when the vaccine would become available and factors that might affect it. For that purpose, they conducted a cross-sectional survey, which revealed that two out of five Greek citizens were not willing or sure about getting a SARS-CoV-2 vaccine, with only 57.7% indicating they would. The authors compared their findings to those of other researchers and concluded that Greeks were more hesitant to get vaccinated against COVID-19 than other Europeans.

3. Research Design

3.1 Methodology

The proposed methodology of our sentiment analysis framework requires an annotated collection of domain-specific tweets that are polarized based on the opinion they express. The vector representation of a given collection of tweets is extracted, and various classifiers are trained to find the best performing one. The goal of this procedure is to create a model capable of predicting the sentiment of future tweets of unknown polarity. Figure 1 illustrates the processes performed to study public opinion towards the COVID-19 vaccination based on social media text.

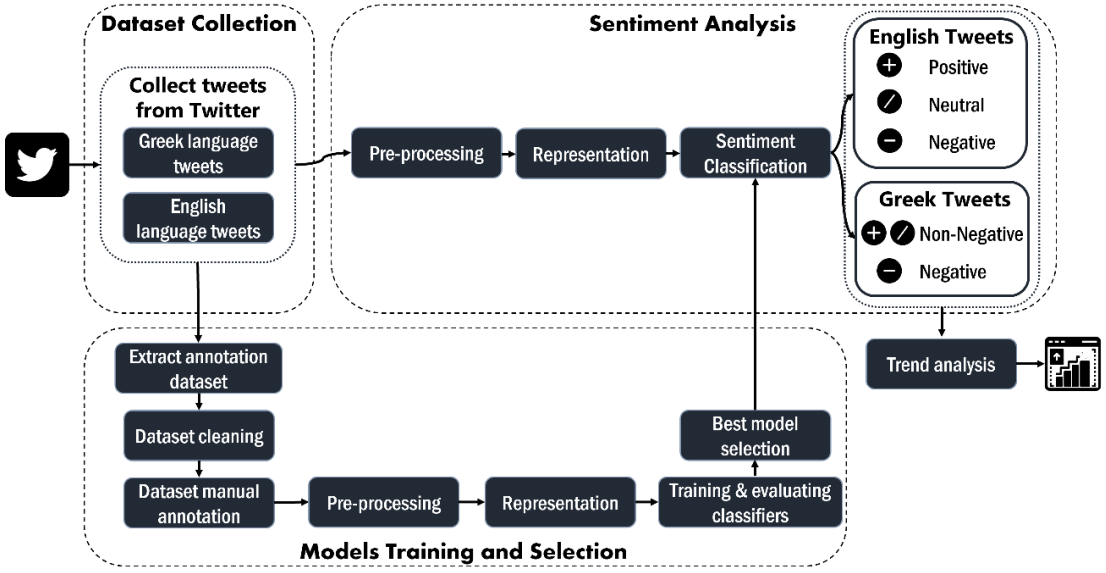

Figure 1: Processes of the proposed sentiment analysis framework.

More in detail, the first step includes the collection of two COVID-19 vaccination datasets, one with English language tweets and one with Greek language tweets. A subset of tweets was then retrieved from each dataset, and each tweet was manually annotated as expressing a positive, neutral, or negative opinion about vaccination. These labelled collections of tweets will be used to train the sentiment analysis models. The next step of our approach is to pre-process and clean the collected tweets because their unstructured and informal nature can degrade the performance of the sentiment detection algorithms.

When it comes to text representation and classification, we chose to examine three approaches: 1) Bag-of-Words representation with classical machine learning, 2) Word Embeddings with classical machine learning and 3) Bidirectional Encoder Representations from Transformers. The performance of various classifiers has been assessed using widely known metrics such as Accuracy, Precision, Recall and F-score. Considering the case of a binary classification for simplicity, i) true positives (TP) denote the real positive observations

that were correctly classified as positive, ii) true negatives (TN) denote the real negative observations that were correctly classified as negative, iii) false positives (FP) are the real negative observations that were incorrectly classified as positive, and lastly, iv) false negatives (FN) are the real positive observations that were incorrectly classified as negative.

Accuracy is the most common evaluation metric, and it represents how frequently the method under consideration made the correct prediction. It is determined by dividing the total number of predictions by the sum of the true predictions. Precision measures how exact the method is, and it is computed as the ratio of correctly predicted positive cases, divided by the total number of predicted positive cases. Recall represents the fraction of the total observations that are correctly predicted by the classifier. Calculating Recall and Precision is usually insufficient. For that reason, F-score is computed that combines both as a weighted average. The definitions of the aforementioned metrics are given in Table 1.

Table 1: Definitions of the metrics used.

Metric Name	**Definition**
Accuracy	$\frac{TP + TN}{TP + FP + FN + TN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F-score	$2 * \frac{Precision * Recall}{Precision + Recall}$

In addition, we visualize the confusion matrix, which is another way of summarizing a classification algorithm's performance. A confusion matrix is a table that contains four fundamental numbers: TP, TN, FP, and FN. The diagonal elements represent instances that have been correctly classified, whereas the off-diagonal components represent instances that have been incorrectly classified. Therefore, the confusion matrix of the best classifier should have only diagonal elements and off-diagonal elements set to zero. Table 2 illustrates an example of a binary classification confusion matrix.

Finally, in the last step of trend analysis, we employed the best performing algorithm to investigate the dynamics of emotional changes towards the vaccination topic over the study period. In greater detail, using the proposed supervised learning model, each tweet is assigned a class label, choosing one among negative, neutral, and positive. The labeled tweets are then analyzed to investigate how public sentiment on Twitter has changed over time, with a focus on local spikes in the daily number of vaccine-related tweets. All the steps taken in our sentiment analysis system are detailed in the subsections that follow.

Table 2: Confusion matrix for binary classification.

		Predicted classes	
		Negative	**Positive**
Actual classes	**Negative**	TN	FP
	Positive	FN	TP

3.2 COVID-19 Vaccine Dataset

In our work, we chose to use a machine learning approach to sentiment analysis, which necessitates the use of a labeled dataset during the training phase of the classification models. To that end, we collected and manually annotated a vaccine-related dataset for each of the languages considered, using the Twitter platform as a data source.

3.2.1 Dataset Collection

The extraction of large-scaled tweet datasets satisfying specific characteristics, such as containing specific keywords, or being tweeted by a specific user, originating from a specific location, is easily accomplished using the Twitter APIs. The retrieved tweets are in JSON format, making it easy to parse them using one of several programming languages. The Twitter API returns metadata, which includes plenty of information such as the date of publishing, the author's username, retweets, hashtags, followers, location, and many more [15]. Hence, we used the Python library Tweepy [53] to fetch Twitter data from the Twitter API. Several COVID-19 vaccine-related hashtags, as listed in Table 3, were used to retrieve tweets. This hashtag collection includes not only covid vaccine-related neutral hashtags (CovidVaccine, COVID19Vaccination), but also covid vaccine-related hashtags that are either positively (GetVaccinated, VaccinesWork, vaccinessaveslives) or negatively (antivaxx, VaccineDeaths, vaccinesdontwork) oriented towards vaccines. As it can be seen, we used almost the same set of hashtags to retrieve tweets in both languages, with one exception: instead of using #vaccinesdontwork, #vaccinessafety, and #vaccinessaveslives we selected the Greek words for vaccination (#εμβολιασμός) and vaccines (#εμβόλια) to retrieve tweets in Greek.

The tweets were gathered over a seven-month period beginning on May 19, 2021 and ending on November 19, 2021.As we already mentioned, we examine both English and Greek components of sentiment analysis. Thus, we collected two distinct Twitter datasets, one for each language. A total of 1,394,535 English language and a total of 61,077 Greek language tweets about the COVID-19 vaccines were identified. The raw tweet collections were reduced

Table 3: Hashtags for tweet search.

Hashtag	**English Language Tweets (N=1,255,554)**	**Greek Language Tweets (N=49,375)**
	Frequency, n	
#vaccinessaveslives	2071	-
#vaccinesafety	634	-
#vaccinesdontwork	578	-
#vaccine	295075	684
#GetVaccinated	260655	37
#CovidVaccine	206812	1434
#vaccinated	181227	266
#vaccination	177481	1803
#VaccinesWork	77371	116
#COVID19Vaccination	24585	54
#vaxxed	14663	4
#antivaxx	6051	6
#vaccinationdone	4365	20
#VaccineDeaths	3986	7
#εμβολιασμος	-	34234
#εμβολια	-	10710

with the goal of removing duplicate tweets, i.e., tweets with the same tweet id that may have been retrieved through different searches, and tweets having different tweet ids but the same tweet text. The cleaned datasets have 1,257,944 and 50,796 English and Greek written tweets, respectively. A summary of the number of unique users, the number of tweets shared by these users, and the number of unique hashtags contained in these tweets are provided in Table 4.

Table 4: Summarization of the cleaned datasets in terms of unique users, tweets, and hashtags.

	English Language Dataset			***Greek Language Dataset***		
	Users	**Tweets**	**Hashtags**	**Users**	**Tweets**	**Hashtags**
Unique	357910	1257944	565253	7310	50796	28614

For each tweet post, we recorded the following features: a) tweet id, a unique 64-bit unsigned integer; b) user screen name; c) number of followers; d) number of user statuses; e) number of retweets; f) full tweet text; g) hashtags; h) date of tweet creation; and i) user location. We also opted not to retrieve or analyze retweets, which are other users' tweets that are re-distributed.

3.2.2 Dataset Annotation

Despite the fact that manual annotation is an expensive and time-consuming task, we chose this approach because we either did not already identify a labeled dataset for sentiment analysis regarding COVID-19 vaccines, like in the case of Modern Greek, or we identified domain-specific datasets that did not perform sufficiently.

In the case of English language tweets, we selected and manually annotated 2403 tweets. A positive label was assigned to tweets containing positive language like expressions of support, positive attitude or emotion, and tweets describing positive situations and events. Accordingly, a negative label was assigned to tweets containing negative language like expressions of judgement, negative attitude, or emotions, as well as tweets describing negative situations and events. A neutral label was assigned to tweets that did not include sentiment terms or phrases, and neither expressed an emotional state. Thus, the neutral category primarily consists of vaccine-related news, as well as tweets containing vaccination-related information or inquiries. It is worth noting that when a tweet included negation before negative or positive phrases, the sentiment polarity was reversed. Consequently, tweets with positive sentiment towards vaccination were assigned the categorical label value 2, tweets with neutral sentiment were assigned the categorical 1abel value 1 and tweets with negative sentiment were assigned the categorical label value 0.

In the case of Greek language tweets, we selected and hand-labelled a small subset of the dataset, consisting of 1424 tweets, to two labels: negative and non-negative to vaccine. Taking into consideration that the number of positive tweets was extremely low, the positive and neutral classes mentioned above were merged into the non-negative one, and the tweets were assigned labels of 0 for negative and 1 for non-negative. However, determining the class to which each tweet belongs is subjective since subtlety expressed opinions may have different interpretations. Table 5 illustrates how the annotated dataset of tweets is distributed among the considered categories in each case. Several examples of the hand-labeled training tweets are given in Table 6.

Table 5: Annotated tweets distribution.

	Category	**Number**	**Total**
English Dataset	Positive	801	2403
	Neutral	801	

	Negative	801	
Greek Dataset	Negative	712	1424
	Non-Negative	712	

Table 6: Examples of hand-labeled tweets.

Label	**Tweet**
Negative (label 0)	Σταματήστε επιτέλους το αναθεματισμένο εμβόλιο astrazeneca στους κάτω των 50 ετών πολίτες.Βλέπετε ότι χάνονται ζωές, παιδια χάνουν τους γονείς τους από την μια μέρα στην άλλη.ΣΤΑΜΑΤΗΣΤΕ ΚΑΙ ΞΥΠΝΗΣΤΕ. #εμβολιασμος #εμβολιο #Μητσοτακης #μητσοτακη_παραιτησου #AstraZeneca
	Μην κάνετε το εμβόλιο θάνατο, θέλουν να μας καταγράψουν όλους !!111εντεκα!11! #εμβολιασμός https://t.co/Su4vGLcMOy
Non-negative (label 1)	Διαβάστε: Η Moderna ζήτησε άδεια για τρίτη δόση από τον FDA https://t.co/N8EEks61HU #Moderna #COVID19 #CovidVaccine #COVIDVaccination
	"Ανεξάρτητα απ' το αν μετράω τις ώρες για τον εμβολιασμό, η οργάνωση φαίνεται !Κύριε @Pierrakakis ακόμη ένα ΜΠΡΑΒΟ" 👏👏👏👏👏👏 #εμβολιασμος https://t.co/o4f5bUT0zS"""

Label	Tweet
Positive (label 2)	I got vaccinated today! Thank you medical science, thank you to the NHS and a massive thank you to the nurses, NHS staff and volunteers giving up their weekends to get us vaccinated! #GetVaccinated #ThankYouNHS https://t.co/9pzraqIUQo
	Delighted to finally have my first vaccine! The whole operation at Whalley Range High School was super efficient, I'm just so grateful to our wonderful NHS staff. #CovidVaccine
Neutral (label 0)	600,000 #kids, ages 12-15 got the #Covid-19 #vaccination in one week according to the #CDC https://t.co/PzrBztF0yB
	The @CDCDirector has confirmed that it is okay to get the #covidvaccine along with other #vaccines. Read full article here https://t.co/RCPUy82AZT https://t.co/T52ONjdDog
Negative (label 1)	Government confirms first death following coronavirus vaccination in India. A 68-year-old man died due to anaphylaxis after taking the vaccine. #COVID19 #CovidVaccine https://t.co/h8wAmXx4Jg
	2 hours at the convention centre and still not in even inside the main area yet for my #vaccine - no wonder this program is a disaster -

	impossible to get a booking but walk-ins are prioritised last #auspol #COVID19Vic

3.2.3 Data Pre-processing

The data collection process is followed by the data pre-processing and cleaning step, which prepares data for further analysis while also ensuring its quality. It has been proven [54] that eliminating URLs, stopwords, and numbers has little effect on classifier performance but is necessary to reduce noise. Experiment results, on the other hand, show that substituting negation and expanding acronyms can enhance the accuracy and f-score of the Twitter sentiment analysis classifier. Thus, the appropriate pre-processing methods are selected for different classifiers of the sentiment classification task at hand.

In our case, the tweet texts were first converted to lowercase using a case-folding operation. The most frequently used emoticons were then replaced with the corresponding words, as it has been shown that they carry useful information and should be considered for sentiment analysis tasks [55]. A number of popular contractions, slang, and informal abbreviations were also replaced with their original forms, and elements like URLs and username mentions were discarded since they did not contribute to our analysis. Numbers and special characters were also eliminated because they provided no relevant information for determining sentiment. Regarding hashtags, we just deleted the hashtag symbol ("#") while keeping the content because they are frequently used in place of normal words and thus store valuable information. The actions outlined above were carried out using Python scripts that leverage regular expressions and the “re” module. For Greek language tweets, we consider an additional step at the start of the cleaning process that includes replacing accented vowels with unaccented ones.

In addition, depending on the classification model being tested, we may or may not eliminate stopwords and conduct lemmatization of the remaining words. Stopwords are words that appear frequently in a corpus but do not provide additional meaning in the analysis, such as articles, conjunctions, or prepositions. Based on the task at hand, we need to eliminate different words from the corpus. The most important words in our case, and in sentiment analysis tasks in general, are those that express sentiment, and eliminating such words would completely change the stance of the text. For example, negative terms like “no” and “don’t”, that are commonly found in stopwords lists should be kept, which is why we modified and used the stopwords list provided by Natural Language Toolkit (NLTK) library [56], after removing any negative words. Following the same rationale, we also used a list of Greek stopwords [57], from which we exclude negative terms that could impact the sentiment analysis. Lemmatization is the process of eliminating a word's inflectional endings and returning it to its base form. In our research, we employ the spacy library [58] to perform lemmatization, either in English or in Greek.

3.3 Text Representation

Before using any text classification algorithm, text data needs to be translated into a numerical feature vector. The simplest method is to employ the Vector Space representation model, either by capturing the presence or absence of each word in the text or by weighting each term based on its importance in the document using Term Frequency-Inverse Document Frequency (TF-IDF). For our first approach, we chose TF-IDF, in which each word is weighted and assigned its respective TF and IDF score. The TF metric determines the number of occurrences of a term in a document, whereas the IDF metric indicates how common or rare a term is across the entire collection of documents. The lower the value, the more frequent the word in the given text, and vice-versa.

The second most common weighted method that has been extensively used in the literature is word embeddings. Word embeddings are vector representations that capture the semantic, morphological, and contextual information of words, allowing words with similar meanings to have similar representations. Several methods have been proposed to derive word embeddings depending on the task at hand. Word2Vec [59], GloVe [60], and FastText [61] are examples of popular word embedding generation models. In this study, we use embeddings generated by the Word2Vec algorithm, which captures the context of words using machine learning techniques such as Recurrent or Deep Neural Networks. To do this, two different learning models exist: the continuous bag-of-words model and the skip-gram model. The former predicts a target word based on the surrounding words, while the latter predicts a set of surrounding words when given a target word.

Word2Vec works on a collection of sentences, building a vocabulary by considering the words that occur in the collection more times than a numeric threshold specified by the user, and then implementing the CBOW or the Skip-gram model to learn a D-dimensional vector representation of each word from the input documents. Word2Vec can be trained using large textual collections such as Wikipedia articles in a specific language or thematic textual corpora, which allow for more accurate tracking of word usage in a given domain. In the model building phase, we have two alternatives: using the sentences derived from the pre-processing phase to learn a new vector representation for each word by utilizing Word2Vec or using a pre-trained Word2Vec model. Afterwards, the vector representation for each sentence in the document is calculated by averaging the vectors of all the words that comprise it. The vector representation of each document is also derived using the same logic. The advantage of this approach over the bag-of-words scheme is that it provides a denser representation for each document.

In addition to learning individual word representations, using Word2Vec, GloVe, or FastText, there is the option of learning word context representations with ELMo, BERT, or GPT2. The superior effectiveness of the latter set of methods over the aforementioned schemes brings them to the forefront of NLP tasks, including text classification. As a result,

in our proposed approach, we chose BERT to test its performance against TF-IDF and Word2Vec.

The following sub-section discusses the results of nine models corresponding to one of the text representation approaches listed above, with the BERT model achieving the best results, for the text classification problem at hand. Therefore, in our framework, BERT is the chosen text representation and classification scheme for both English and Greek tasks.

3.4 Learning Algorithms

In order to successfully evaluate the sentiment about vaccination in the collected tweets, a machine learning approach has been followed. According to relevant field surveys [62], [63], there are numerous classification algorithms that may be trained to yield binary or multiclass text classification results. Using the annotated dataset as training data, the performance of the following prominent different classifiers was reviewed in our research: Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and Extreme Gradient Boosting (XGBoost). Following that, we go through some additional information about each classifier.

a) Support Vector Machine

Support Vector Machines (SVM) are supervised machine learning methods that are commonly used for tasks such as classification, regression, and outlier detection. This algorithm aims to determine a function in a multi-dimensional space that can optimally separate data with known class labels [64]. One of its advantages is that it allows for the specification of different kernel functions for the decision function, including linear, polynomial, or even a custom kernel [65]. In this study, we used SVM with a linear kernel.

b) Logistic Regression

Logistic Regression (LR) is another supervised machine learning technique that is often used for categorical binary classification problems. It can, however, be modified so as to model multiple classes of events [66]. Logistic Regression is a simple classification algorithm that works well with linearly separable classes [67]. Thus, it has been used in a wide range of applications, including spam detection, speech recognition and other.

c) Random Forest

Random Forest (RF) classifier is an ensemble of decision trees, that is widely employed for classification. Each tree in the RF is supplied with the input vector and votes for a class. The individual tree votes are then aggregated and the class with the most votes is selected as the model's final decision. Consequently, this machine learning algorithm is easily developed and implemented, and produces robust classification results [68], [69], [70].

d) XGBoost Algorithm

Gradient boosting decision trees are a type of ensemble machine learning models that is commonly used for classification and regression tasks. The eXtreme gradient boosting tree

(XGBoost) is a more advanced version of the gradient boosting model, capable of handling sparse data and missing values, as well as avoiding overfitting. Moreover, in terms of classification accuracy and computation speed, XGBoost outperforms other tree boosting algorithms [71].

e) Bidirectional Encoder Representations from Transformers

BERT stands for Bidirectional Encoder Representations from Transformers, and it is a transformer-based language model that has been pretrained using a large corpus of unlabeled text from the English Wikipedia and BooksCorpus [72]. This text representation technique developed by Google combines various deep learning techniques including bidirectional encoder LSTM and Transformers [73]. In comparison to previous language models, BERT is unique in that it can read text input in both directions at once, rather than only left-to-right or right-to-left. Using its pretraining as a base layer, we may fine-tune BERT to a user's specifications with a single additional output layer, producing cutting-edge models for a variety of applications [74].

The paper [72] describes two model sizes for BERT, $BERT_{BASE}$ and $BERT_{LARGE}$. In this research, we selected $BERT_{BASE}$[1] model with L = 12 layers, a hidden size H = 768, A = 12 self-attention heads and 110M parameters. In contrast, $BERT_{LARGE}$ has L = 24 layers, a hidden size H = 1024, A = 16 self-attention heads and 340M parameters. There is also a Modern Greek version of BERT language model [75] called GreekBERT[2], which is trained on the Greek parts of 1) Wikipedia, 2) the European Parliament Proceedings Parallel Corpus, and 3) OSCAR [76]. The authors have published a model similar to $BERT_{BASE}$, that has derived state-of-the-art results in numerous NLP tasks, and can be fine-tuned for domain-specific problems [75].

[1] https://huggingface.co/bert-base-uncased
[2] https://huggingface.co/nlpaueb/bert-base-greek-uncased-v1

4. Experimental Results

In order to perform the experiments in Python programming language, we used libraries such as scikit-learn[3], gensim[4] and xgboost[5] for classical machine learning models, while the BERT model was implemented using the transformers[6] and torch[7] libraries. The considered models and their results are detailed in the following.

We used 5-fold cross validation technique to evaluate the considered models as well as grid search approach to determine the optimal hyperparameters of each classifier and increase its performance. During the k-fold cross validation procedure the training set is divided into k smaller subsets, k-1 of which are used to train the model while the remaining part is used to validate the resulting model [77]. Grid search performs an exhaustive search through a user specified collection of an algorithms' hyperparameters [78].

First, we used a TF-IDF Vectorizer with the n-gram parameter set to (1,2), which refers to unigrams and bigrams. Moving on to the next approach, which is word embeddings with classical machine learning, recent studies either employ publicly available pre-trained Word2Vec model, such as the ones discussed above, or trained Word2Vec models on domain-specific text collections. In our case, we chose to train a Word2Vec model on our collected training data to obtain vector representations for all the unique words present in the corpus, instead of using one of the existing pre-trained word-vectors. The vector representation of each tweet is derived by taking the mean of all the word vectors present in the tweet. The Word2Vec model we trained uses the skip-gram training algorithm and the dimensionality of its word vectors equals to 100.

Additionally, as already mentioned, we evaluated the selected algorithms by considering whether the elimination of stopwords and the implementation of lemmatization improved the performance of the chosen algorithms. We present the findings of the scenario that most classifiers perform best in. For the TF-IDF scheme, this means performing both stopwords elimination and lemmatization in each tweet. For the English language Word2Vec model, we only performed stopwords removal, while for the Greek language Word2Vec model we did not apply either stopwords removal nor lemmatization.

The last approach we explored is the BERT language model. The optimal values for the hyperparameters of BERT are task-specific, but Devlin et al. [72] suggest the following possible values regarding the batch sizes (16, 32), the learning rate (5e-5, 3e-5, 2e-5) and number of epochs (2, 3, 4). We further trained two existing language models, $\text{BERT}_{\text{BASE}}$ and GreekBERT, on our collected English and Greek language social media corpus respectively,

[3] https://scikit-learn.org/stable/
[4] https://pypi.org/project/gensim/
[5] https://xgboost.ai/
[6] https://huggingface.co/docs/transformers/index
[7] https://pypi.org/project/torch/

finetuning them using a batch size of 16, a learning rate of 2e-5 and 4 epochs. Tables 7 and 8 show the results achieved by each machine learning classifier after applying the 5-fold cross validation and grid search techniques. In both tables, the best accuracy values per representation technique and classifier are in bold.

Table 7: Classification Performance Analysis Table (English language datasets).

Representation Technique	Classifier	Class	Precision	Recall	F-score	Accuracy
TF-IDF	SVM (C=2, kernel='linear')	*Positive*	0.86	0.85	0.86	0.84
		Neutral	0.79	0.80	0.80	
		Negative	0.88	0.87	0.87	
	LR (C=10)	*Positive*	0.88	0.85	0.86	**0.85**
		Neutral	0.79	0.84	0.81	
		Negative	0.89	0.87	0.88	
	RF (max_features='log2', min_samples_split=4, n_estimators=200)	*Positive*	0.82	0.81	0.82	0.79
		Neutral	0.68	0.85	0.76	
		Negative	0.89	0.71	0.79	
	GBM (gamma=0.1, max_depth=3, n_estimators=200)	*Positive*	0.78	0.79	0.79	0.77
		Neutral	0.67	0.82	0.74	
		Negative	0.88	0.70	0.78	
Word2Vec	SVM (C=10, gamma=0.1)	*Positive*	0.89	0.87	0.88	**0.86**
		Neutral	0.81	0.83	0.82	
		Negative	0.88	0.88	0.88	
	LR (C=5, max_iter=500)	*Positive*	0.89	0.89	0.89	0.85
		Neutral	0.81	0.80	0.81	
		Negative	0.87	0.87	0.87	
	RF (n_estimators=300)	*Positive*	0.86	0.85	0.85	0.82
		Neutral	0.75	0.82	0.79	
		Negative	0.86	0.80	0.83	
	GBM (gamma=0.01, max_depth=9, n_estimators=200)	*Positive*	0.85	0.87	0.86	0.83
		Neutral	0.79	0.81	0.80	
		Negative	0.85	0.82	0.84	

BERT	*Positive*	0.90	0.90	0.90	**0.91**
	Neutral	0.91	0.90	0.90	
	Negative	0.93	0.95	0.94	

4.1 English language data

When employing the TF-IDF technique for text representation, we can see that Logistic Regression is the best performing classifier, followed by the Support Vector Machine model, with 85% and 84% accuracy scores, respectively. On the other hand, the Random Forest, and Extreme Gradient Boosting algorithms with 79% and 77% accuracy each, are the worst performing classifiers.

The results reported in Table 7 show that the use of word embeddings improves the performance of all algorithms except Logistic Regression, for which the accuracy score remains unchanged. The Random Forest and Extreme Gradient Boosting classifiers showed the greatest improvement in accuracy with 82% and 83% scores, respectively. The Support Vector Machine model, on the other hand, outperforms all models with an accuracy of 86%, and it is the best classifier when word embeddings are used as a text representation strategy.

Having an accuracy of 91% the $BERT_{BASE}$ model exceeds the performance of all the aforementioned classifiers. We also observe that for both weighting schemes, TF-IDF and word embeddings, the considered algorithms perform less well in terms of precision, recall and f-score, for the neutral class, which does not hold true for the BERT language model.

4.2 Greek language data

Despite the fact that Modern Greek is an under-resourced language, the TF-IDF weighting method employed with classical machine learning produces results that are comparable to the English models, as shown in Table 8. The Support Vector Machine is the best performing classifier, with an accuracy of 86%, followed closely by Logistic Regression, which has an accuracy of 85%. The other two algorithms, Random Forest and Extreme Gradient Boosting performed poorly with accuracies of 78% and 76%, respectively.

When it comes to word embeddings, the performance of Random Forest and Extreme Gradient Boosting classifiers improves by 4% for the former and 3% for the latter one. It is worth noting that the accuracy of the SVM and LR models has reduced to 83% and 81%, respectively. It is possible that this is due to the fact that embedding approaches are incapable of dealing with unfamiliar words or words that do not belong to the vocabulary.

Despite this decrease in accuracy, the Support Vector Machine model outperformed the other classifiers, this time being followed by the RF model. As a consequence, we can conclude that the Support Vector Machine and Logistic Regression algorithms work best when using

the TF-IDF weighting scheme, whilst the Random Forest and Extreme Gradient Boosting classifiers perform better when using the word embeddings technique.

The best performance among all the explored approaches is achieved when finetuning the GreekBERT language model with an accuracy of 93%. This model also outperforms the others in terms of precision, recall, and f-score across all classes, negative and non-negative.

Table 8: Classification Performance Analysis Table (Greek language datasets).

Representation Technique	**Classifier**	**Class**	**Precision**	**Recall**	**F-score**	**Accuracy**
TF-IDF	SVM (C=1, kernel='linear')	*Non-Negative*	0.90	0.83	0.86	**0.86**
		Negative	0.82	0.89	0.85	
	LR (C=10)	*Non-Negative*	0.89	0.82	0.86	0.85
		Negative	0.81	0.89	0.85	
	RF (max_features='log2', n_estimators=500)	*Non-Negative*	0.75	0.90	0.82	0.78
		Negative	0.84	0.65	0.73	
	GBM (gamma=0.1, max_depth=3, n_estimators=100)	*Non-Negative*	0.75	0.84	0.79	0.76
		Negative	0.78	0.66	0.72	
Word2Vec	SVM (C=5, gamma=0.5)	*Non-Negative*	0.88	0.80	0.84	**0.83**
		Negative	0.79	0.87	0.83	
	LR (C=10,max_iter=500)	*Non-Negative*	0.86	0.79	0.82	0.81
		Negative	0.77	0.85	0.81	
	RF (min_samples_split=5,n_estimators=500)	*Non-Negative*	0.86	0.81	0.83	0.82
		Negative	0.79	0.84	0.81	
	GBM (gamma=0.1, max_depth=9, n_estimators=100)	*Non-Negative*	0.82	0.79	0.80	0.79
		Negative	0.76	0.79	0.78	

BERT	*Non-Negative*	0.94	0.91	0.93	**0.93**
	Negative	0.92	0.95	0.93	

To summarize, most classifiers trained on the Greek language dataset outperformed or were comparable to those trained on the English language dataset. However, the fine-tuned BERT models have the highest accuracy scores for both languages. Figure 2 depicts the confusion matrix for each model. In the following section, we employed the top performing classifier for each language, $BERT_{BASE}$ and GreekBERT, to examine the tweets gathered throughout the specified time period.

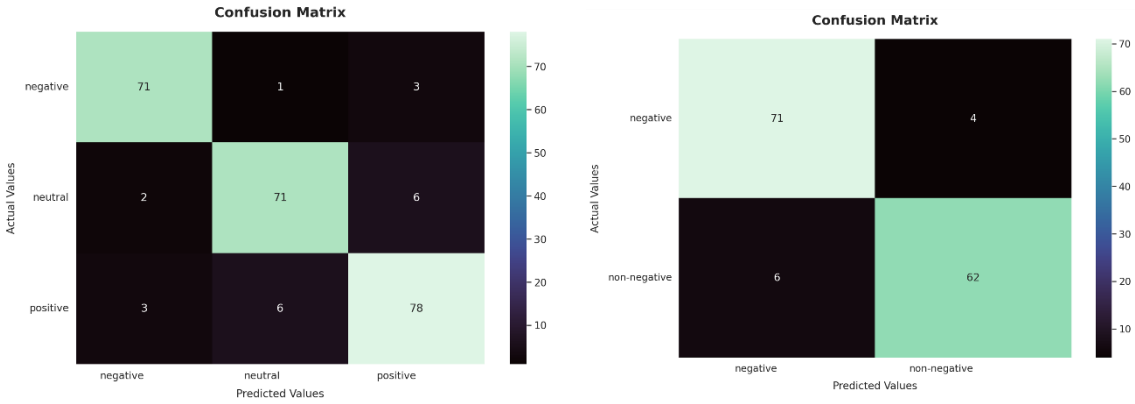

Figure 2: The confusion matrices for $BERT_{BASE}$ (left) and GreekBERT (right).

The next two sections describe the final process of our proposed research design, trend analysis. Python, along with its wordninja, matplotlib, and plotly libraries, were used to generate the results and figures that follow.

4.3 English language dataset trend analysis

After discarding the training tweets, we applied the finetuned $BERT_{BASE}$ model to the entire 1,255,554 tweets in the English language corpus, in order to perform sentiment analysis and determine the polarity of each tweet and classify it based on three labels, namely: positive, negative, and neutral. Then, we analyze the classified tweets over time to determine sentiment changes of the public towards the vaccination topic. Such variations may occur in response to known or unknown social, vaccine-related events. Several types of visualizations are used to demonstrate the dynamics of sentiment evolution in the collected corpus.

Figures 3 and 4 show the daily distribution of the retrieved vaccine-related tweets, and the number of them shared per month, respectively. In the first figure, we can see that the number of tweets fluctuates, constantly increasing or decreasing. The lowest number of tweets was observed on October 31, while the highest amount was noticed on July 30, when about 10,000 tweets were retrieved. One of the conclusions that can be drawn from Figure 4 is that the

number of tweets gathered is rather high across all of the examined months, despite the fact that May and November are half-studied.

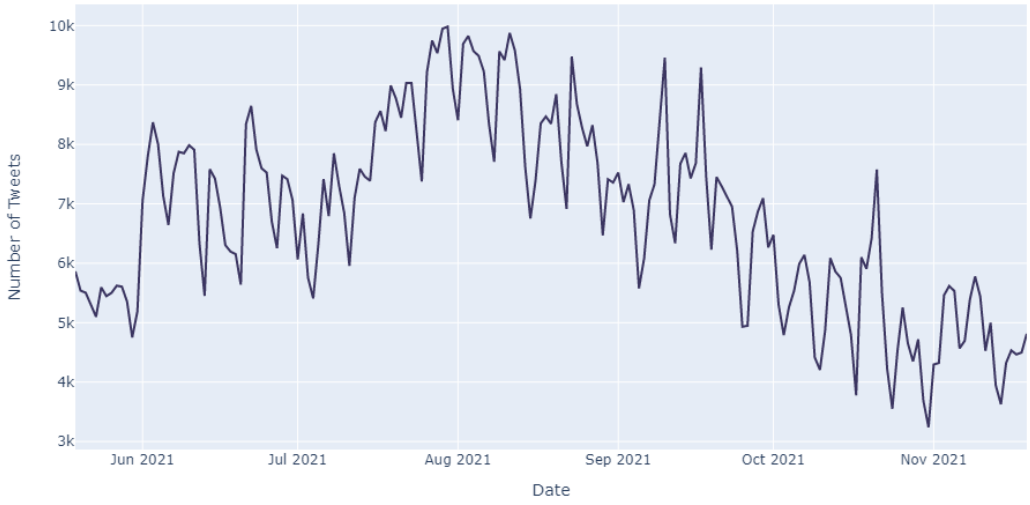

Figure 3: Number of collected tweets on a daily basis.

The total number of tweets distribution by percentage based on the different sentiment categories, is illustrated in the left side of Figure 5. As can be seen, the 'Neutral' category has the highest percentage of tweets, reaching 39.9%, followed by the 'Negative' category with 36.5% and the 'Positive' category with 23.6%. Having the majority of tweets fall into the neutral category is a common social network behavior that has also been reported in previous study [16]. More interestingly, similar trends can be observed on the right side of Figure 5, but only for the first three months of the study period, where the percentage of tweets with neutral sentiment outnumbers those with negative and positive sentiment. We also notice that in May and June the percentage of tweets belonging to positive sentiment category exceeds that of negative although it is continually declining until November when there is a slight increase. In July, negative tweets were more than positive tweets, while beginning in August and continuing through the end of the analysis period, the percentage of tweets with negative sentiment exceeds the percentages of neutral and positive tweets.

Since the above is an aggregated result, which may conceal fluctuations that occurred within the time range, we further perform a daily analysis. Hence, we plot a daily timeline showing the evolution of vaccine-related tweets based on sentiment, between May 19, 2021 and November 19, 2021 (Figure 6). As expected, the predominant sentiment is neutral until July 15, 2021, when negative sentiment starts to dominate, with a few exceptions. The number of positive tweets does not exhibit dramatic fluctuations except for September 17, 2021, and October 21, 2021, when it peaks.

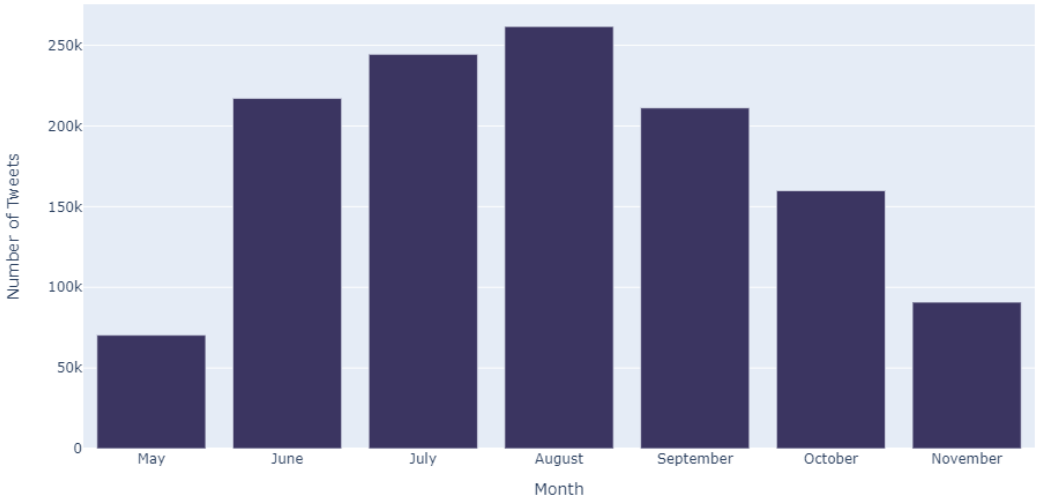

Figure 4: Number of the retrieved tweets shared each month.

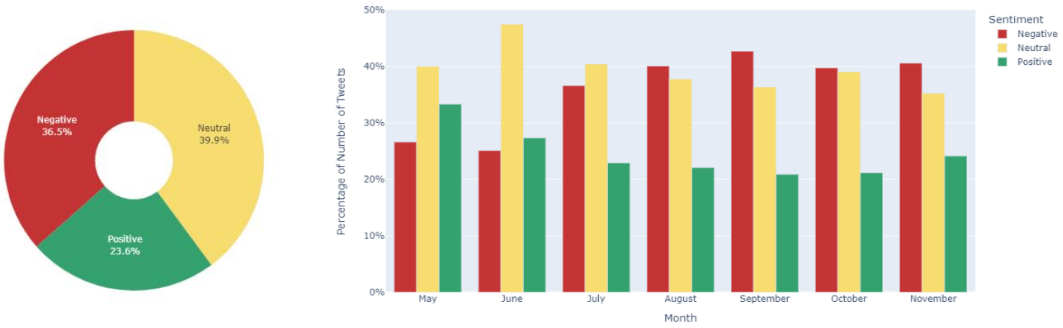

Figure 5: Tweets percentage distribution by sentiment category collectively (left) and for each month individually (right).

Tweets expressing negative sentiments are linked to a variety of concerns, but most of them focus on antivaccine conspiracy theories being spread resulting in more COVID-19 victims, vaccine safety and delta variant concerns, reporting side effects, reaction to the antivaccine movement and health restrictions. Tweets expressing positive sentiment, on the other hand, are mostly about trusting science, urging people to get vaccinated, thanking people who volunteer on vaccination programs as well as people who choose to vaccinate, and being happy about the loosening of restrictions that widespread vaccination results in. The neutral sentiments are related to health reports and studies, as well as news and information regarding vaccination locations, open slots, and dose administration updates. Those conclusions are based on a manual inspection of random tweets from each sentiment, but they are also validated by the following word clouds analysis.

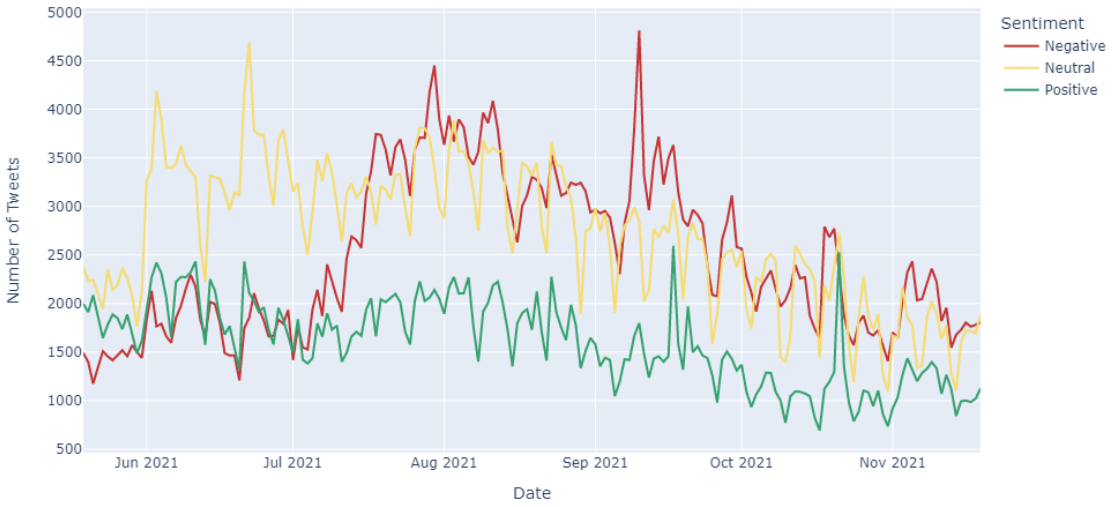

Figure 6: Time series for the collected tweets on a daily basis based on sentiment.

A word cloud is a graphical representation of commonly used words, in which the size of the word in the representation is determined by the popularity and frequency of the word. Thus, the more times a word appears in the text collection, the bigger its size in the cloud. Word clouds are effective at detecting the most frequently discussed topics and can be implemented in a variety of settings [79]. In this study, we created word clouds of 100 most frequently appearing words associated with each sentiment category to gain more insights about COVID-19 vaccine-related tweets. Originally, we attempted to generate word clouds based on frequency of word occurrence, but the resulting visualizations provided little useful information because nearly the identical words appeared in all three negative, neutral, and positive word clouds. The log-likelihood values were then used to generate word clouds with more relevant information for vaccine-related opinions. Figure 7 shows word clouds colored by sentiment, with red, yellow, and green representing negative, neutral, and positive, respectively.

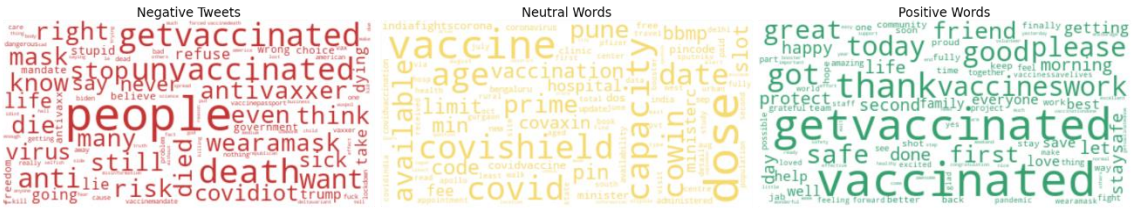

Figure 7: Word clouds showing the most frequent appearing words in vaccine-related tweets for each sentiment category. The left image is formed from negative tweets, the middle image is formed from neutral tweets and the right image is formed from positive tweets.

According to Figure 7, some of the most frequently appearing words in negative tweets are people, death, unvaccinated, covidiot and antivaxxer, indicating that many Twitter users

expressed their negative thoughts about people who refused to receive the vaccine. Frequent terms such as mask, mandate, virus, sick and risk, reveal that people are still concerned about the consequences of the coronavirus and the mask mandates. Secondary words in the negative word cloud include stupid, refuse, wrong, government, and trump.

Words that appear often in neutral tweets include “vaccine”, “dose”, “covid”, “capacity”, “available”, and “date”, indicating a need for information on the vaccination program. “Covishield” and “cowin” are other common terms, with the former referring to the Indian version of the AstraZeneca vaccine and the latter to the Indian online vaccination site. If we look closely, we can uncover secondary words like “bbmp”, which is an Indian Municipal Corporation, “indiafightscorona”, a popular hashtag used during the pandemic, “pune”, “an Indian city”, “slot”, “prime”, “minister”, “hospital”, “limit” and “covaxin”, Indias’ indigenous COVID-19 vaccine. These terms suggest that a significant number of tweets is originating from Indian users.

Finally, when overviewing the rightmost subplot of Figure 7, we witness terms like “vaccinated”, “thank”, “great”, “please”, which indicate that people are getting vaccinated, supporting the vaccines, and urging more people to vaccinate with hashtags like “getvaccinated” and “vaccineswork”. Some more related words shown in a smaller size in the word cloud are “safe”, “done”, “happy”, “first”, “second”, “love”, “grateful”, “protect” and the hashtag “staysafe”.

We notice that many of the most frequently appearing words in the tweets collection are hashtags, indicating that are widely used by Twitter users. The most tweeted hashtags for the time period under consideration are listed in Figure 8. As shown, the majority of them are related to the SARSCoV-2 vaccine, with two, #WearAMask and #StaySafe referring to the compliance with protective measures following immunization.

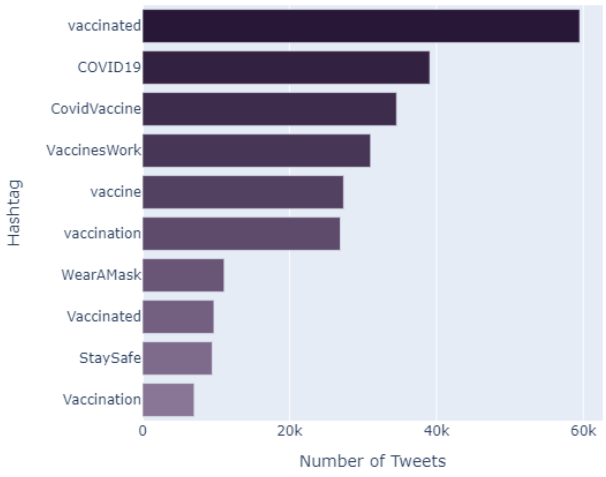

Figure 8: Top-10 most tweeted hashtags.

4.4 Greek language dataset trend analysis

Following the same rationale as with English written tweets, we removed the training tweets from the complete set of tweets and applied the finetuned GreekBERT model to all 49,375 tweets in the Greek language corpus. The objective of this process was to determine the sentiment label of each tweet and classify it based on two categories, namely: negative and non-negative. The classified tweets are then analyzed over time to investigate sentiment changes among Greek Twitter users regarding vaccines. Such changes in sentiment may be attributed to known or unknown social, vaccine-related events. We used several types of visualizations to highlight the dynamics of sentiment evolution in the collected corpus.

Figure 9 shows the distribution of vaccine-related tweets written in Greek, over the 7-month study period. We notice that the number of tweets retrieved is relatively low with small increases or decreases over time, except for the period between June 28, 2021, and August 7, 2021, when the highest and lowest peaks respectively are recorded. The quantity of tweets during this time fluctuates dramatically, while another significant peak is observed on August 24, 2021. The massive growth in the number of tweets on June 28, 2021 was in response to the prime minister's announcement about the administration of a pre-paid card as an incentive for those aged 18-25 to get vaccinated against COVID-19 [80]. Twitter users made fun of the idea, complained about not getting a similar reward due to their age, and mainly accused the government of bribing young people.

Figure 10 shows the total number of COVID-19 vaccine-related tweets shared each month. The majority of the tweets (17,354) are retrieved during July, with the number sharply decreasing in the following months.

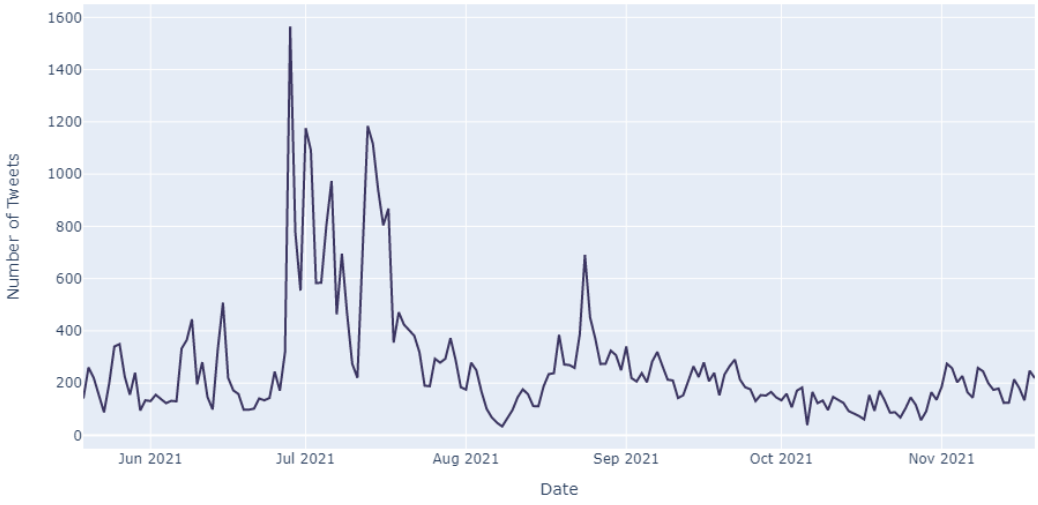

Figure 9: Tweet distribution during the considered period.

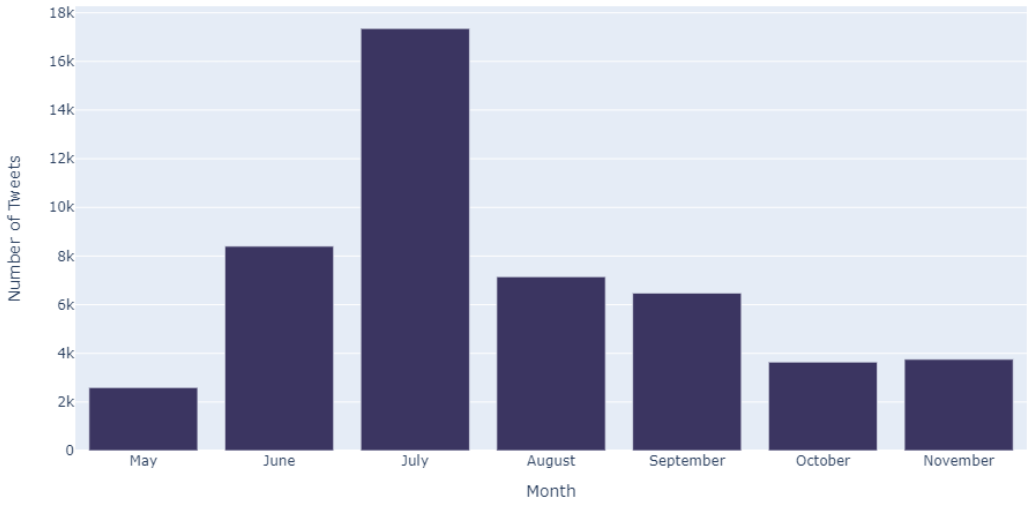

Figure 10: Number of the retrieved tweets shared each month.

The tweets distribution of the Greek language dataset reveals that the majority of the tweets (60.1%) belong to the ‘Negative’ category, while the remaining 39.9% of tweets belong to the 'Non-negative' category. As shown in the right part of Figure 11, tweets with negative sentiment outnumber tweets with non-negative sentiment in all months except May, where the percentage of non-negative tweets exceeds that of negative tweets by 6%.

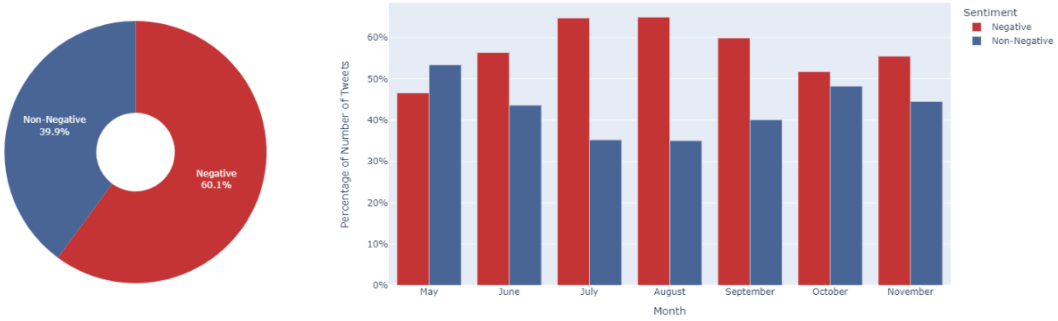

Figure 11: Tweets percentage distribution by sentiment category collectively (left) and for each month individually (right).

Next, we study the time series of the tweets based on sentiment, between May 19, 2021 and November 19, 2021. Figure 12 shows the daily progression of negative and non-negative sentiments. As expected, the predominant sentiment is negative, but its trend is closely followed by the non-negative trend, so there are days when they coincide and a few when tweets of non-negative sentiment outnumber tweets of negative sentiment. Negative sentiments are focused on vaccine deaths and severe side effects, beliefs that the vaccines are

dangerous and poisonous and that the pandemic is not real, dissatisfaction with government's pandemic policies, and frustration with people who firmly refuse to get vaccinated. On the other hand, non-negative sentiments are mostly about vaccination news and pandemic updates from Greece and throughout the world, latest vaccine studies, details about variants of concern, and the third dose of the vaccine. Finally, there are a few non-negative tweets praising Greece's well-organized vaccination system.

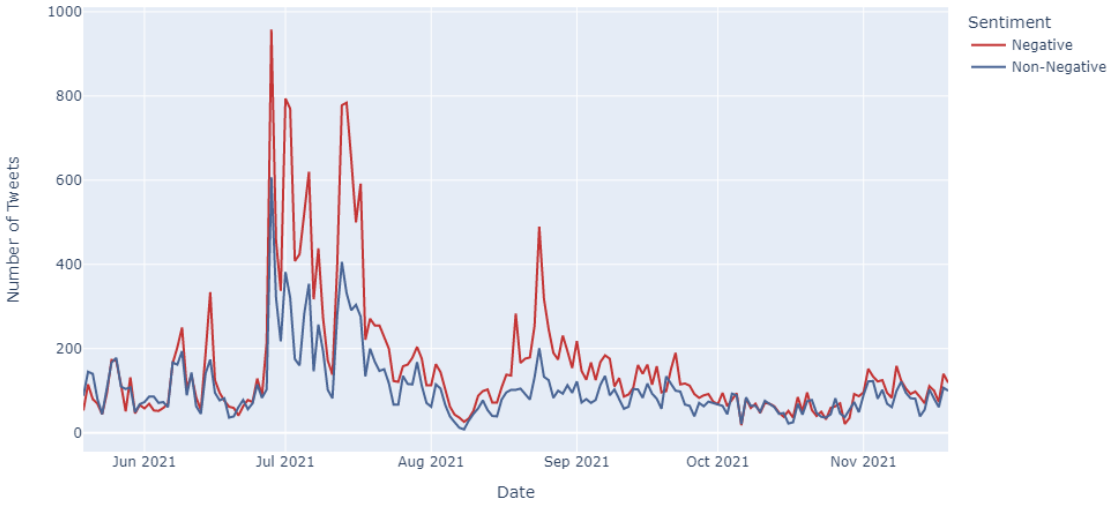

Figure 12: Tweet distribution on a daily basis based on sentiment.

Following that, we created word clouds for each sentiment category using the log-likelihood values, as shown in Figure 13. The red word cloud represents negative tweets, whereas the blue word cloud represents non-negative tweets. The most interesting findings were identified in the negative category. The word cloud is dominated by references to Greece's prime minister and the government in general, the mass media, as well as rude and inelegant characterizations of them. Twitter users, in particular, commonly used phrases like circus and carnival to describe the government and the prime minister's pandemic-related actions, accusing them of forcing a Greek junta and demanding their resignation. Other frequent negative words include vaccine deniers, vaccine side-effects, coronavirus variants and deaths. Among the keywords associated with non-negative sentiment, the most popular ones are the Greek translations of coronavirus, pandemic, vaccines, covid cases and mutations, third, second, dose, children, delta, referring to the delta variant spreading among the population, and hashtags including "covidgr", "covidgreece", and "rollingupsleeves".

Finally, we conduct a hashtag analysis by identifying the ones that appeared frequently. The most tweeted hashtags, as listed in the Figure 14, are largely related to the coronavirus and COVID-19 vaccine. However, there is one (#κρουσματα) referring to covid cases and another one (#ανεμβολιαστοι) referring to unvaccinated people.

Figure 13: Word clouds showing the most frequent appearing words in vaccine-related tweets for each sentiment category. The left image is formed from negative tweets, while the right image is formed from non-negative tweets.

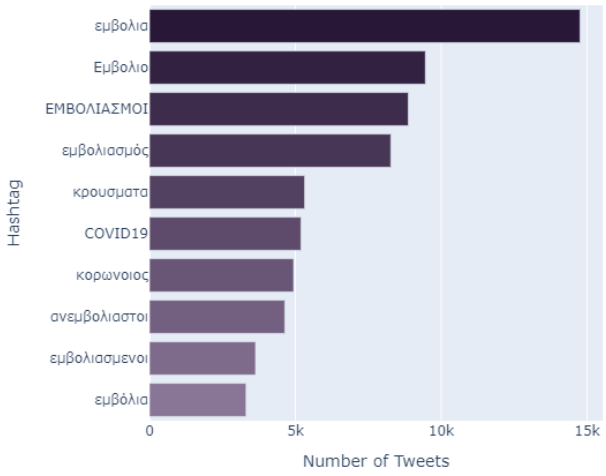

Figure 14: Top-10 most tweeted hashtags.

5. Conclusions

5.1 Discussion

This study identifies and examines state-of-the-art mining approaches and applications on social media text that target English and Modern Greek. The objective was to track the evolution of sentiment regarding COVID-19 vaccines over the 7-month period between May 19, 2021 and November 19, 2021. This was accomplished by creating a machine learning based sentiment analysis model based on pre-annotated tweets, that is capable of analyzing and classifying the entire user-generated Twitter post in both languages. Several classical machine learning classifiers and language models were examined for both English and Greek language, with their accuracy rates ranging from 76% to 93%. The best performing classifier for each language was identified by four performance metrics. The proposed framework classifies English language tweets into three main classes: negative, neutral, and positive, employing BERT with 91% accuracy. Greek language tweets are classified into two classes, namely negative and non-negative, employing GreekBERT with an accuracy of 93%.

When the English dataset was studied collectively, the prevailing sentiment was neutral, but at the daily level, neutral was only predominant during the first three months. Except for a few days, the general sentiment has been negative for the months that followed. When we compare the percentage of tweets belonging to each sentiment category at the start of the period, May 19, to the end of the period, November 19, we find that negative sentiment increased by nearly 12%, positive sentiment decreased by 11%, and the percentage of neutral sentiment remained almost the same. The evolution of all three sentiments is marked by a series of spikes, some larger than others.

Negative sentiment also dominates the Greek dataset both overall and on a daily basis, with the exception of a few days where non-negative sentiment outnumbers negative sentiment. When the percentage of tweets on May 19 and November 19 was compared, it was found that negative tweets grew by 17% while non-negative tweets declined by 17%. A number of spikes is observed in the evolution of negative sentiment tweets, which is closely followed by the evolution of non-negative sentiment tweets. The sudden increase in the number of Greek tweets on the day of a prime minister's announcement, is proof that policies and events entrain sentiment variations evident in the time series of tweets.

Early detection of sentiment shifts can be highly useful now that most nations across the world are vaccinating their populations with booster doses and contemplating new preventive measures. This would enable public healthcare organizations to develop effective crisis management strategies, better inform the public with accurate and reliable news, alleviate disease-specific concerns, minimize the distribution of fake news, control panic, and boost vaccination acceptance.

A sentiment analysis task based on social media data can yield useful insights regarding online discussions around COVID-19 vaccines, which governments can use to improve reaction times and plan ahead of time to prevent risks influenced by social media. According to the findings of this study Twitter is an efficient platform for understanding public opinions, concerns, awareness, and emotions as well as targeting and evaluating health initiatives. Twitter users primarily discuss and react to public health issues and interventions providing information that government agencies can use to determine whether health messages are resonating.

Given the novelty of the disease and the unprecedent speed with which vaccines were developed there was an exacerbation in safety and efficacy concerns regarding COVID-19 vaccines. As vaccination process is still hampered by several barriers and new cases and deaths are exponentially growing there is an urgent need to monitor vaccine-related conversations on social media. Our work demonstrates that an easily implemented method may be used to monitor vaccine sentiment in near real-time, allowing us to identify events that influence it on a global or national scale. The advantage is that policymakers and health experts know what people are thinking about vaccines, regulations around them, and the disease in general at any given time, allowing them to respond immediately to possible concerns. Such knowledge may be useful in taking the required actions and planning interventions to increase vaccine uptake and reduce vaccine hesitancy.

The pandemic has upended people's lives all around the world, placing millions of individuals at risk of infection or increased mental health issues, like fear, sadness, worry, and anxiety. However, quick responses and actions facilitated by social media analysis, aimed at minimizing and preventing negative emotional and psychological impacts will enhance global health and well-being amid crises such as the SARS-CoV-2 pandemic.

5.2 Limitations and Future Work

This study has a number of limitations. To begin, we collected Twitter data by querying for a limited set of hashtags related to COVID-19 vaccines, for each language. This set may have been incomplete, as new trending hashtags have emerged over time to keep Twitter topics grouped. Another restriction stems from the question of whether our findings and Twitter users are representative of the general Greek-speaking and English-speaking public.

On the one hand, Twitter data is a valuable source for studying real-time social media user-generated content about coronavirus vaccines, yet Twitter users do not represent the entire population, and their tweets do not constitute a representative sample of public opinion. They simply represent netizens' views and emotions towards vaccination. Additional difficulties arise for Greek tweets because Twitter's reach in Greece is smaller than in other countries, reducing its representative power even further.

The next constraint is linked to data collection and processing. It would be interesting to have a larger dataset, especially for Modern Greek, as well as data over a longer timeframe to

analyze sentiment evolution. Social media data is typically noisy in nature, having different types and formatting styles. However, it is difficult to completely filter noise and eliminate irrelevant tweets, especially when dealing with huge corpora. Moreover, the selection of data processing techniques such as stop words removal, lemmatization, etc. introduces subjectivity, which may have a significant impact on the research's outcomes.

Additionally, it is questionable whether a tweet's sentiment content could be considered truthful and reliable. This introduces a bias into the results of our analysis necessitating cautious interpretation. This is deteriorated by the possibility of confounding events. For example, the announcement of the freedom pass administration, triggered a steep rise in negative sentiment in our analysis. The unfavorable feeling around this event was directed at the government's policy rather than the vaccine itself.

Finally, our proposed framework is based on employing machine learning techniques to evaluate tweets. While machine learning can process large amounts of data considerably faster than human approaches, it has the disadvantage of not performing as well as human inspection, particularly in tweets containing irony and sarcasm. Future research objectives could involve building better performing sentiment classification algorithms.

There exists an overwhelming amount of information about COVID-19 vaccines available online, making it difficult to determine the optimal information sources. Our study simply analyzes Twitter data to understand people's sentiment, but future research could conduct the same analysis with a larger social media dataset, acquired from multiple social network sites including Facebook. Moreover, data sources other than social media could be employed in the analysis.

Another interesting future direction includes the emotion detection of tweets and their classification into different emotions, such as happiness, fear, sadness, trust, anger and so on. This is known as emotion analysis, and it has the potential to improve citizens' and society's understanding. Further work should consider studying the geographic distribution of tweets and their sentiment, comparing how sentiment changes across different countries around the world and correlating significant spikes in sentiment to major social events.

Bibliography

[1] "Timeline: WHO's COVID-19 response", *Who.int*, 2021. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline#event-72. [Accessed: 27- Dec- 2021].

[2] "WHO Coronavirus (COVID-19) Dashboard", *Covid19.who.int*, 2021. [Online]. Available: https://covid19.who.int/. [Accessed: 27- Dec- 2021].

[3] M. Ni et al., "Mental Health, Risk Factors, and Social Media Use During the COVID-19 Epidemic and Cordon Sanitaire Among the Community and Health Professionals in Wuhan, China: Cross-Sectional Survey", *JMIR Mental Health*, vol. 7, no. 5, p. e19009, 2020. Available: 10.2196/19009 [Accessed 27 December 2021].

[4] "Topic: Social media use during coronavirus (COVID-19) worldwide", *Statista*, 2021. [Online]. Available: https://www.statista.com/topics/7863/social-media-use-during-coronavirus-covid-19-worldwide/. [Accessed: 27- Dec- 2021].

[5] R. Merchant and N. Lurie, "Social Media and Emergency Preparedness in Response to Novel Coronavirus", *JAMA*, vol. 323, no. 20, p. 2011, 2020. Available: 10.1001/jama.2020.4469 [Accessed 27 December 2021].

[6] "COVID-19 Vaccine Visualization", *Covid-19vaccinetracker.org*, 2021. [Online]. Available: https://www.covid-19vaccinetracker.org/#Top-of-Page. [Accessed: 27- Dec- 2021].

[7] "Ten health issues WHO will tackle this year", *Who.int*, 2021. [Online]. Available: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019. [Accessed: 27- Dec- 2021].

[8] "Twitter: most users by country | Statista", *Statista*, 2021. [Online]. Available: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/. [Accessed: 27- Dec- 2021].

[9] H. Rosenberg, S. Syed and S. Rezaie, "The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic", *CJEM*, vol. 22, no. 4, pp. 418-421, 2020. Available: 10.1017/cem.2020.361 [Accessed 27 December 2021].

[10] E. Cambria, B. Schuller, Y. Xia and C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis", *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15-21, 2013. Available: 10.1109/mis.2013.30 [Accessed 27 December 2021].

[11] X. Wang, C. Zou, Z. Xie and D. Li, "Public Opinions towards COVID-19 in California and New York on Twitter", 2020. Available: 10.1101/2020.07.12.20151936 [Accessed 27 December 2021].

[12] B. Liu, "Sentiment Analysis and Opinion Mining", *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1-167, 2012. Available: 10.2200/s00416ed1v01y201204hlt016 [Accessed 27 December 2021].

[13] D. Chatzakou and A. Vakali, "Harvesting Opinions and Emotions from Social Media Textual Resources", *IEEE Internet Computing*, vol. 19, no. 4, pp. 46-50, 2015. Available: 10.1109/mic.2015.28 [Accessed 27 December 2021].

[14] R. Feldman, "Techniques and applications for sentiment analysis", *Communications of the ACM*, vol. 56, no. 4, pp. 82-89, 2013. Available: 10.1145/2436256.2436274 [Accessed 27 December 2021].

[15] A. Giachanou and F. Crestani, "Like It or Not", *ACM Computing Surveys*, vol. 49, no. 2, pp. 1-41, 2016. Available: 10.1145/2938640 [Accessed 27 December 2021].

[16] E. D'Andrea, P. Ducange, A. Bechini, A. Renda and F. Marcelloni, "Monitoring the public opinion about the vaccination topic from tweets analysis", *Expert Systems with Applications*, vol. 116, pp. 209-226, 2019. Available: 10.1016/j.eswa.2018.09.009 [Accessed 27 December 2021].

[17] P. Alexander and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", 2010.

[18] M. Mansoor, K. Gurumurthy, A. R U and V. Badri Prasad, "Global Sentiment Analysis Of COVID-19 Tweets Over Time", *CoRR*, 2020. [Accessed 27 December 2021].

[19] M. Hung et al., "Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence", *Journal of Medical Internet Research*, vol. 22, no. 8, p. e22590, 2020. Available: 10.2196/22590 [Accessed 27 December 2021].

[20] C. Shofiya and S. Abidi, "Sentiment Analysis on COVID-19-Related Social Distancing in Canada Using Twitter Data", *International Journal of Environmental Research and Public Health*, vol. 18, no. 11, p. 5993, 2021. Available: 10.3390/ijerph18115993 [Accessed 27 December 2021].

[21] S. Boon-Itt and Y. Skunkan, "Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study", *JMIR Public Health and Surveillance*, vol. 6, no. 4, p. e21978, 2020. Available: 10.2196/21978 [Accessed 27 December 2021].

[22] S. Kaur, P. Kaul and P. Zadeh, "Monitoring the Dynamics of Emotions during COVID-19 Using Twitter Data", *Procedia Computer Science*, vol. 177, pp. 423-430, 2020. Available: 10.1016/j.procs.2020.10.056 [Accessed 27 December 2021].

[23] J. Samuel, G. Ali, M. Rahman, E. Esawi and Y. Samuel, "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification", *Information*, vol. 11, no. 6, p. 314, 2020. Available: 10.3390/info11060314 [Accessed 27 December 2021].

[24] D. Kaila and D. Prasad, "Informational Flow on Twitter – Corona Virus Outbreak – Topic Modelling Approach", *International Journal of Advanced Research in Engineering and Technology (IJARET)*, vol. 11, no. 3, pp. 128-134, 2020. [Accessed 27 December 2021].

[25] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag and A. Hassanien, "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media", *Applied Soft Computing*, vol. 97, p. 106754, 2020. Available: 10.1016/j.asoc.2020.106754 [Accessed 27 December 2021].

[26] X. Yuan and A. Crooks, "Examining Online Vaccination Discussion and Communities in Twitter", *Proceedings of the 9th International Conference on Social Media and Society*, 2018. Available: 10.1145/3217804.3217912 [Accessed 27 December 2021].

[27] V. Raghupathi, J. Ren and W. Raghupathi, "Studying Public Perception about Vaccination: A Sentiment Analysis of Tweets", *International Journal of Environmental Research and Public Health*, vol. 17, no. 10, p. 3464, 2020. Available: 10.3390/ijerph17103464 [Accessed 27 December 2021].

[28] J. Du, J. Xu, H. Song, X. Liu and C. Tao, "Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets", *Journal of Biomedical Semantics*, vol. 8, no. 1, 2017. Available: 10.1186/s13326-017-0120-6 [Accessed 27 December 2021].

[29] J. Du, J. Xu, H. Song and C. Tao, "Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data", *BMC Medical Informatics and Decision Making*, vol. 17, no. 2, 2017. Available: 10.1186/s12911-017-0469-6 [Accessed 27 December 2021].

[30] T. Hu et al., "Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective", *Journal of Medical Internet Research*, vol. 23, no. 9, p. e30854, 2021. Available: 10.2196/30854 [Accessed 27 December 2021].

[31] J. Lyu, E. Han and G. Luli, "COVID-19 Vaccine–Related Discussion on Twitter: Topic Modeling and Sentiment Analysis", *Journal of Medical Internet Research*, vol. 23, no. 6, p. e24435, 2021. Available: 10.2196/24435 [Accessed 27 December 2021].

[32] L. Cotfas, C. Delcea, I. Roxin, C. Ioanas, D. Gherai and F. Tajariol, "The Longest Month: Analyzing COVID-19 Vaccination Opinions Dynamics From Tweets in the Month Following the First Vaccine Announcement", *IEEE Access*, vol. 9, pp. 33203-33223, 2021. Available: 10.1109/access.2021.3059821 [Accessed 27 December 2021].

[33] P. Pristiyono, M. Ritonga, M. Ihsan, A. Anjar and F. Rambe, "Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm", *IOP Conference Series: Materials Science and Engineering*, vol. 1088, no. 1, p. 012045, 2021. Available: 10.1088/1757-899x/1088/1/012045 [Accessed 27 December 2021].

[34] C. Villavicencio, J. Macrohon, X. Inbaraj, J. Jeng and J. Hsieh, "Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes", *Information*, vol. 12, no. 5, p. 204, 2021. Available: 10.3390/info12050204 [Accessed 27 December 2021].

[35] A. Khakharia, V. Shah and P. Gupta, "Sentiment Analysis of COVID-19 Vaccine Tweets Using Machine Learning", *SSRN Electronic Journal*, 2021. Available: 10.2139/ssrn.3869531 [Accessed 27 December 2021].

[36] A. Hussain et al., "Artificial intelligence-enabled analysis of UK and US public attitudes on Facebook and Twitter towards COVID-19 vaccinations", 2020. Available: 10.1101/2020.12.08.20246231 [Accessed 27 December 2021].

[37] N. Sattar and S. Arifuzzaman, "COVID-19 Vaccination Awareness and Aftermath: Public Sentiment Analysis on Twitter Data and Vaccinated Population Prediction in the USA", *Applied Sciences*, vol. 11, no. 13, p. 6128, 2021. Available: 10.3390/app11136128 [Accessed 27 December 2021].

[38] S. Yousefinaghani, R. Dara, S. Mubareka, A. Papadopoulos and S. Sharif, "An analysis of COVID-19 vaccine sentiments and opinions on Twitter", *International Journal of Infectious Diseases*, vol. 108, pp. 256-262, 2021. Available: 10.1016/j.ijid.2021.05.059 [Accessed 27 December 2021].

[39] R. Marcec and R. Likic, "Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines", *Postgraduate Medical Journal*, 2021. Available: 10.1136/postgradmedj-2021-140685 [Accessed 27 December 2021].

[40] V. Athanasiou and M. Maragoudakis, "A Novel, Gradient Boosting Framework for Sentiment Analysis in Languages where NLP Resources Are Not Plentiful: A Case Study for Modern Greek", *Algorithms*, vol. 10, no. 1, p. 34, 2017. Available: 10.3390/a10010034 [Accessed 27 December 2021].

[41] A. Braoudaki, E. Kanellou, C. Kozanitis and P. Fatourou, "Hybrid Data Driven and Rule Based Sentiment Analysis on Greek Text", *Procedia Computer Science*, vol. 178, pp. 234-243, 2020. Available: 10.1016/j.procs.2020.11.025 [Accessed 27 December 2021].

[42] G. Markopoulos, G. Mikros, A. Iliadi and M. Liontos, "Sentiment Analysis of Hotel Reviews in Greek: A Comparison of Unigram Features", *Cultural Tourism in a Digital Era*, pp. 373-383, 2015. Available: 10.1007/978-3-319-15859-4_31 [Accessed 27 December 2021].

[43] N. Spatiotis, I. Mporas, M. Paraskevas and I. Perikos, "Sentiment Analysis for the Greek Language", *Proceedings of the 20th Pan-Hellenic Conference on Informatics*, 2016. Available: 10.1145/3003733.3003769 [Accessed 27 December 2021].

[44] M. Giatsoglou, M. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis and K. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings", *Expert Systems with Applications*, vol. 69, pp. 214-224, 2017. Available: 10.1016/j.eswa.2016.10.043 [Accessed 27 December 2021].

[45] G. Alexandridis, I. Varlamis, K. Korovesis, G. Caridakis and P. Tsantilas, "A Survey on Sentiment Analysis and Opinion Mining in Greek Social Media", *Information*, vol. 12, no. 8, p. 331, 2021. Available: 10.3390/info12080331 [Accessed 27 December 2021].

[46] G. Kalamatianos, D. Mallis, S. Symeonidis and A. Arampatzis, "Sentiment analysis of greek tweets and hashtags using a sentiment lexicon", *Proceedings of the 19th Panhellenic Conference on Informatics*, 2015. Available: 10.1145/2801948.2802010 [Accessed 27 December 2021].

[47] A. Tsakalidis et al., "Building and evaluating resources for sentiment analysis in the Greek language", *Language Resources and Evaluation*, vol. 52, no. 4, pp. 1021-1044, 2018. Available: 10.1007/s10579-018-9420-4 [Accessed 27 December 2021].

[48] D. Beleveslis, C. Tjortjis, D. Psaradelis and D. Nikoglou, "A Hybrid Method for Sentiment Analysis of Election Related Tweets", *2019 4th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*, 2019. Available: 10.1109/seeda-cecnsm.2019.8908289 [Accessed 27 December 2021].

[49] D. Antonakaki, D. Spiliotopoulos, C. V. Samaras, P. Pratikakis, S. Ioannidis and P. Fragopoulou, "Social media analysis during political turbulence", *PLOS ONE*, vol. 12, no. 10, 2017. Available: 10.1371/journal.pone.0186836 [Accessed 27 December 2021].

[50] N. Makrynioti and V. Vassalos, "Sentiment Extraction from Tweets: Multilingual Challenges", *Big Data Analytics and Knowledge Discovery*, pp. 136-148, 2015. Available: 10.1007/978-3-319-22729-0_11 [Accessed 27 December 2021].

[51] D. Kydros, M. Argyropoulou and V. Vrana, "A Content and Sentiment Analysis of Greek Tweets during the Pandemic", *Sustainability*, vol. 13, no. 11, 2021. Available: 10.3390/su13116150 [Accessed 27 December 2021].

[52] G. Kourlaba et al., "Willingness of Greek general population to get a COVID-19 vaccine", *Global Health Research and Policy*, vol. 6, no. 1, 2021. Available: 10.1186/s41256-021-00188-1 [Accessed 27 December 2021].

[53] "Tweepy", *Tweepy.org*, 2021. [Online]. Available: https://www.tweepy.org/. [Accessed: 27- Dec- 2021].

[54] Z. Jianqiang and G. Xiaolin, "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis", *IEEE Access*, vol. 5, pp. 2870-2879, 2017. Available: 10.1109/access.2017.2672677 [Accessed 27 December 2021].

[55] A. Debnath, N. Pinnaparaju, M. Shrivastava, V. Varma and I. Augenstein, "Semantic Textual Similarity of Sentences with Emojis", *Companion Proceedings of the Web Conference 2020*, 2020. Available: 10.1145/3366424.3383758 [Accessed 27 December 2021].

[56] S. Bird, E. Klein and E. Loper, *Natural language processing with Python*. Bejing: O'Reilly, 2009.

[57] "GitHub - stopwords-iso/stopwords-el: Greek stopwords collection", *GitHub*, 2021. [Online]. Available: https://github.com/stopwords-iso/stopwords-el. [Accessed: 27- Dec- 2021].

[58] *Spacy.io*, 2021. [Online]. Available: https://spacy.io/api/lemmatizer. [Accessed: 27- Dec- 2021].

[59] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality", *Advances in Neural Information Processing Systems*, vol. 26, 2013. [Accessed 27 December 2021].

[60] J. Pennington, R. Socher and C. Manning, "Glove: Global Vectors for Word Representation", *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014. Available: 10.3115/v1/d14-1162 [Accessed 27 December 2021].

[61] P. Bojanowski, E. Grave, A. Joulin and T. Mikolov, "Enriching Word Vectors with Subword Information", *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135-146, 2017. Available: 10.1162/tacl_a_00051 [Accessed 27 December 2021].

[62] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes and D. Brown, "Text Classification Algorithms: A Survey", *Information*, vol. 10, no. 4, 2019. Available: 10.3390/info10040150 [Accessed 27 December 2021].

[63] V. Vijayan, K. Bindu and L. Parameswaran, "A comprehensive study of text classification algorithms", *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017. Available: 10.1109/icacci.2017.8125990 [Accessed 27 December 2021].

[64] J. Dukart, "Basic Concepts of Image Classification Algorithms Applied to Study Neurodegenerative Diseases", *Brain Mapping*, pp. 641-646, 2015. Available: 10.1016/b978-0-12-397025-1.00072-5 [Accessed 27 December 2021].

[65] "1.4. Support Vector Machines", *scikit-learn*, 2021. [Online]. Available: https://scikit-learn.org/stable/modules/svm.html. [Accessed: 27- Dec- 2021].

[66] A. Bartosik and H. Whittingham, "Evaluating safety and toxicity", *The Era of Artificial Intelligence, Machine Learning, and Data Science in the Pharmaceutical Industry*, pp. 119-137, 2021. Available: 10.1016/b978-0-12-820045-2.00008-8 [Accessed 27 December 2021].

[67] A. Subasi, "Machine learning techniques", *Practical Machine Learning for Data Analysis Using Python*, pp. 91-202, 2020. Available: 10.1016/b978-0-12-821379-7.00003-5 [Accessed 27 December 2021].

[68] S. Misra and Y. Wu, "Machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales with feature extraction and feature ranking", *Machine Learning for Subsurface Characterization*, pp. 289-314, 2020. Available: 10.1016/b978-0-12-817736-5.00010-7 [Accessed 27 December 2021].

[69] M. Mushtaq and A. Mellouk, "Methodologies for Subjective Video Streaming QoE Assessment", *Quality of Experience Paradigm in Multimedia Services*, pp. 27-57, 2017. Available: 10.1016/b978-1-78548-109-3.50002-3 [Accessed 27 December 2021].

[70] S. Misra and H. Li, "Noninvasive fracture characterization based on the classification of sonic wave travel times", *Machine Learning for Subsurface Characterization*, pp. 243-287, 2020. Available: 10.1016/b978-0-12-817736-5.00009-0 [Accessed 27 December 2021].

[71] Y. Chang, K. Chang and G. Wu, "Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions", *Applied Soft Computing*, vol. 73, pp. 914-920, 2018. Available: 10.1016/j.asoc.2018.09.029 [Accessed 27 December 2021].

[72] J. Devlin, M. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", *Proceedings of the 2019 Conference of the North*, pp. 4171–4186, 2019. Available: 10.18653/v1/n19-1423 [Accessed 27 December 2021].

[73] A. Imran, S. Daudpota, Z. Kastrati and R. Batra, "Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets", *IEEE Access*, vol. 8, pp. 181074-181090, 2020. Available: 10.1109/access.2020.3027350 [Accessed 27 December 2021].

[74] "What is BERT (Language Model) and How Does It Work?", *SearchEnterpriseAI*, 2021. [Online]. Available: https://searchenterpriseai.techtarget.com/definition/BERT-language-model. [Accessed: 27- Dec- 2021].

[75] J. Koutsikakis, I. Chalkidis, P. Malakasiotis and I. Androutsopoulos, "GREEK-BERT: The Greeks visiting Sesame Street", *11th Hellenic Conference on Artificial Intelligence*, 2020. Available: 10.1145/3411408.3411440 [Accessed 27 December 2021].

[76] B. Sagot, "OSCAR", *OSCAR*, 2021. [Online]. Available: https://oscar-corpus.com/. [Accessed: 27- Dec- 2021].

[77] "3.1. Cross-validation: evaluating estimator performance", *scikit-learn*, 2021. [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html. [Accessed: 27- Dec- 2021].

[78] S. Chan and P. Treleaven, "Continuous Model Selection for Large-Scale Recommender Systems", *Handbook of Statistics*, pp. 107-124, 2015. Available: 10.1016/b978-0-444-63492-4.00005-8 [Accessed 27 December 2021].

[79] C. Conner, J. Samuel, A. Kretinin, Y. Samuel and L. Nadeau, "A Picture for the Words! Textual Visualization in Big Data Analytics", 2019. Available: 10.13140/RG.2.2.25351.83360 [Accessed 27 December 2021].

[80] PM Mitsotakis presents 'Freedom Pass' for 18-25 year olds that get vaccinated. Capital.gr. (2021). Retrieved 27 December 2021, from https://www.capital.gr/english/3555769/pm-mitsotakis-presents-freedom-pass-for-18-25-year-olds-that-get-vaccinated.