Investigating the Sentiment in Italian Long-COVID Narrations

Maria Chiara Martinis[0000−0002−4868−7767]
Ileana Scarpino*[000−0002−5216−968X]
Chiara Zucco[0000−0003−0048−0457]
Mario Cannataro[0000−0003−1502−2387]

Data Analytics Research Center, Department of Medical and Surgical Sciences,
University Magna Graecia of Catanzaro, Italy
{martinis,ileana.scarpino,chiara.zucco,cannataro}@unicz.it

Abstract. Through an overview of the history of the disease, Narrative Medicine (NM) aims to define and implement an effective, appropriate and shared treatment path. In the context of COVID-19, several blogs were produced, among those the "Sindrome Post COVID-19" contains narratives related to the COVID-19 pandemic. In the present study, different analysis techniques were applied to a dataset extracted from such "Sindrome Post COVID-19" blog. The first step of the analysis was to test the VADER polarity extraction tool. Then the analysis was extended through the application of Topic Modeling, using Latent Dirichlet Allocation (LDA). The results were compared to verify the correlations between the polarity score obtained through VADER and the extracted topics through LDA. The results showed a predominantly negative polarity consistent with the mostly negative topics represented by words on post virus symptoms. The results obtained derive from three different approaches applied to the COVID narrative dataset. The first part of the analysis corresponds to polarity extraction using the VADER software, where, from the score, polarity was inferred by dichotomizing the overall score. In the second part, topic modeling through LDA was applied, extracting a number of topics equal to three. The third phase is based on the objective of finding a qualitative relationship between the polarity extracted with VADER and the latent topics with LDA, considering it a semi-supervised problem. In the end, the presence of polarized topics was explored and thus a correspondence between sentiment and topic was found.

Keywords: Sentiment Analysis · Polarity Detection · Narrative Medicine · VADER · Topic Modeling · Latent Dirichlet Allocation

1 Introduction

The COVID-19 pandemic had a strong impact not only on people's physical health, but also on their mental health [6], [21], [11]. Social networks were widely

* Martinis and Scarpino contributed equally to this work.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65

used during the isolation period as they allowed people to tell about their experience in dealing with the pandemic, sharing their physical and emotional state.

The large amount of data available for collection and analysis has therefore allowed the scientific community to apply Natural Language Processing (NLP) techniques, specifically sentiment analysis, to synthesize and investigate various aspects related to the COVID-19 pandemic narrative.

There are several studies in the literature on topic extraction in tweets related to COVID-19 pandemic [30],[4], [2]. In particular, the study of Medford et al. [14] collected tweets related to COVID-19 in January 2020, during the initial phase of the outbreak, measuring the themes and frequency of keywords related to infection prevention practices, identifying the polarity of sentiment and predominant emotions in the tweeting and conducting a topic modeling of the topics of discussion over time through Latent Dirichlet Allocation.

Sentiment analysis techniques [32] were applied to identify topics and trends in Twitter users' sentiment about the pandemic. The authors' conclusions assert that the use of the social network has allowed activating sometimes coping mechanism to combat feelings of isolation related to long-term social distancing [29] and sometimes exacerbating concern and disinformation [20]. The evaluation of the most discussed topics by people who have dealt with COVID-19 disease can lead to field applications of text mining techniques useful in various fields, e.g. polarity analysis, which involves a classification into three labels: positive, negative and neutral, is useful to identify the depth of emotions expressed by users.

In Natural Language Processing, the polarity detection task falls under the umbrella of automatic text classification. Typically, the approaches for polarity classification are divided into two macro-categories: the lexicon-based approaches and the machine learning-based approaches [16]. Lexicon-based approaches start from the assumption that the overall polarity of the text unit can be inferred from the polarity of the words found in the text unit, using a set of predefined rules. Machine learning-based approaches aim to classify the polarity of a text by training a machine learning algorithm with a set of already tagged examples [32]. An example of such a tool is VADER, a lexicon-based sentiment analysis tool specifically adapted to sentiments expressed in social media, which also works well on texts from other domains.[7]

Mathematical and computer models are applied in research to analyze communications about health problems. Detecting emotions from users' reflections is useful for the purpose of diagnosis of diseases and other medical emergencies or epidemics [25], [9], [27]. Particular interest is given to the narratives of patients with Post-Acute Sequelae of COVID-19 (PASC). Post-Acute Sequelae of COVID-19, also known as "post-COVID-19 syndrome" or "Long-COVID," refers to those convalescent individuals in whom prolonged and often debilitating sequelae persist [15]. The clinical symptomatology includes several main manifestations, often affecting different organ systems. In addition to fatigue, malaise and dyspnoea, PASC patients may also be affected by a number of psychiatric disorders, including depression, anxiety, and post-traumatic stress disorder [26].

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65

Some studies found evidence of mood and cognitive impairment that urgently requires the development of targeted therapies and telemedicine support [1]. In others, prevalence estimation was conducted to identify PASC and predict fluctuations in the number of people with persistent symptoms over time [18].

With the huge popularity of social media platforms, humans express their thoughts and feelings more openly than ever before, therefore sentiment analysis has quickly become a hot research field that may help monitor and understand opinions and emotions in different application domains, as for instance marketing and financial analysis, customer services, recommendation systems, but also clinical applications [10].

The present work focuses on the application of Natural Language Processing (NLP) and Sentiment Analysis (SA) methods, in particular polarity detection VADER tool to extract sentiment from Narrative Medicine textual sources containing narratives of subjects with post-covid syndrome (PASC). In a previous work [24], the main focus was investigation of Topic Modeling Techniques to extract meaningful insights in Italian Long COVID narrations. The automatic collection was performed through web scraping of NM Italian blogs, i.e. "Sindrome Post COVID-19" [1] that produced a dataset composed of 73 narratives related to PASC.

The rest of the paper is organized as follows. In Section 2 the background information about NM and NLP is introduced, and an overview of polarity detection and topic modeling application on COVID-19 related text is discussed. Section 3 describes the data collection procedure, the dataset used and the analysis methodology. Section 4 discusses the result of the analysis. Finally, Section 5 concludes the paper and outlines future works.

2 Background on sentiment analysis of COVID-19 texts

Narrative Medicine is a field of medicine that helps all professionals in the health care system to carefully accommodate the experiences of people living with a disease and their caregivers through research and clinical practice. To analyze written and oral narratives, e.g. extracting polarity from texts and other relevant information, polarity detection and topic modeling techniques are adopted. Sentiment Analysis is an emerging methodology that contributes to the understanding of human emotions from social media reviews. In [12] the authors conducted the global sentiment analysis of tweets related to coronavirus and how the sentiment of people in different countries varied over time. Twitter is one of the most commonly used social media platforms, and it showed a massive increase in tweets related to coronavirus, including positive, negative and neutral tweets. Early detection of COVID-19 sentiment from collected tweets allowed for better understanding and management of the pandemic.
In another study [8], researchers analyzed emotions extracted from tweets for sentiment classification using various feature sets and classifiers. Tweets were

[1] https://www.sindromepostcovid19.it/

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65

classified into positive, negative and neutral sentiment classes. The performance of machine learning (ML) and deep learning (DL) classifiers is evaluated using different evaluation metrics (accuracy, precision, recall and F1 score).
In [28] the tweets were analyzed to understand the mood of citizens and students during the COVID-19 pandemic. Sentiment analysis was conducted in relation to the age of users, and the results showed that the entity of tweets was higher in young people during the pandemic.
In [31] authors reviewed the Twitter opinion of nine states of the United States (U.S.) on COVID-19 from April 1 to April 5, 2020 by developing the most popular methods of machine learning and deep learning to predict users' sentiments toward COVID-19 based on tweets. In the work, coding techniques were compared with two separate word embedding systems Word2vec and Glo2vec. Opinions on the introduction of the COVID-19 vaccine varied among people; some showed concern about safety, others supported vaccination as an effective mitigation strategy.
In [19], researchers examined the level of hesitation of the COVID-19 vaccine as part of the effort to combat the infectious virus. The pros and cons of vaccination have become a key topic on social media platforms. The results show that hesitation toward the COVID-19 vaccine is gradually decreasing over time, suggesting that society's positive views of vaccination have gradually increased.

In the context of COVID-19, storytelling is an additional resource for dealing with past and present events during the pandemic, as it provides concrete tools to health care professionals to support health prevention and treatment.

Long COVID is not only a new disease, but also a disease that, perhaps unique in recent history, emerged largely without the patient's clinician acting as a witness or sounding board. Since COVID-19 turns out to be an occupational disease in that it has disproportionately affected health care workers, a socio-narratological approach is necessary in the Long COVID field since both patient trust and obligation on the part of physicians come into play. The story of Long COVID as it unfolded in 2020, includes an overarching metanarrative of absent listeners: the collective failure – arguably for good reasons – of clinicians to acknowledge, interpret or act on their patients' stories and plights [22].
In [23][24] standard Text Mining (TM) techniques were applied for topic modeling to characterize narrative medicine texts written on COVID-19. To extract polarity in texts written in English, the original VADER software was used, a version that was extended to the Italian language, thus creating VADER-IT [13].

3 Materials and Methods

The aim of the present study is to sentiment analysis and text mining approaches for the analysis of a blog about the COVID-19 disease, named "Sindrome post COVID-19"[2], containing narrative of patients affected by Long-COVID (in short Long-COVID testimonies). The dataset of "Sindrome post COVID-19" can be

[2] https://www.sindromepostcovid19.it/

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65

considered an important attempt to create a database to collect the testimonies of those who have had, or are having, PASC syndrome. People who shared their experiences were asked to provide specific information about their symptoms after the test was negative.

3.1 Long-COVID PASC Dataset

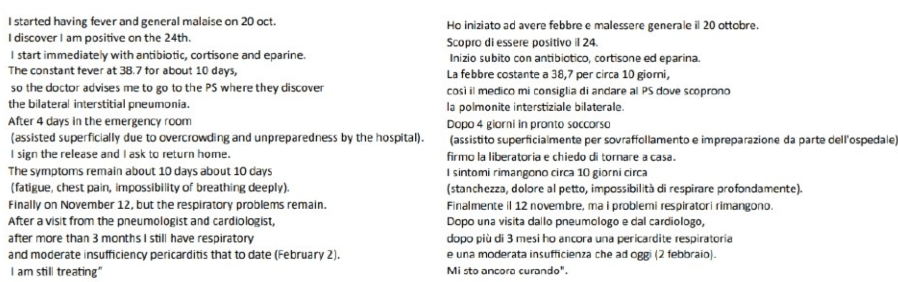
In a recent work [24], we analyzed a dataset extracted by "Sindrome Post COVID-19" containing 73 narrative texts to Post-acute Sequelae of COVID-19 (PASC), also known as Long-COVID syndrome.
For the present study the texts were automatically collected and translated by API Google into English to allow the use of original VADER.

Both data collection and data analysis were performed in Python, a high-level programming language. Python's "Beautiful Soap" and "requests" libraries have been used to automatically collect the testimonies and to store the data in a single CSV file.

For instance, to collect data from the “Sindrome post COVID-19" blog, the Python code:

- collects all the URL of the testimonies per page,
- connects to the page and, by parsing the HTML document:
  - extracts only the textual narration,
  - performs a basic text cleaning,
  - tags it as "PASC",
- the extracted information is stored in a CSV file format.

The described process is then iterated through all the narrations posted on the blog. The information is stored in a *CSV* text file for the subsequent analysis, storing the data sorted by chronological order from the most recent to the oldest post.
Fig. 1 shows an example of text extracted from the dataset.



I started having fever and general malaise on 20 oct.
I discover I am positive on the 24th.
I start immediately with antibiotic, cortisone and eparine.
The constant fever at 38.7 for about 10 days,
so the doctor advises me to go to the PS where they discover
the bilateral interstitial pneumonia.
After 4 days in the emergency room
(assisted superficially due to overcrowding and unpreparedness by the hospital).
I sign the release and I ask to return home.
The symptoms remain about 10 days about 10 days
(fatigue, chest pain, impossibility of breathing deeply).
Finally on November 12, but the respiratory problems remain.
After a visit from the pneumologist and cardiologist,
after more than 3 months I still have respiratory
and moderate insufficiency pericarditis that to date (February 2).
I am still treating"

Ho iniziato ad avere febbre e malessere generale il 20 ottobre.
Scopro di essere positivo il 24.
Inizio subito con antibiotico, cortisone ed eparina.
La febbre costante a 38,7 per circa 10 giorni,
così il medico mi consiglia di andare al PS dove scoprono
la polmonite interstiziale bilaterale.
Dopo 4 giorni in pronto soccorso
(assistito superficialmente per sovraffollamento e impreparazione da parte dell'ospedale)
firmo la liberatoria e chiedo di tornare a casa.
I sintomi rimangono circa 10 giorni circa
(stanchezza, dolore al petto, impossibilità di respirare profondamente).
Finalmente il 12 novembre, ma i problemi respiratori rimangono.
Dopo una visita dallo pneumologo e dal cardiologo,
dopo più di 3 mesi ho ancora una pericardite respiratoria
e una moderata insufficienza che ad oggi (2 febbraio).
Mi sto ancora curando".

Fig. 1. Example of extracted text fragment.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65

3.2 Sentiment Analysis using VADER

Sentiment analysis is useful for detecting positive, negative or neutral sentiment in text. It is often used by companies to detect sentiment in social data, assess brand reputation, and understand customers. Several pre-trained language models have also been proposed in the biomedical field, such as BioBert [4], ClinicalBert [5], Medicines [5], calBert [5], MedBert [6], and PubMedBert [7].
A standard approach to build pre-trained language models in the biomedical field consists of training a general BERT model on biomedical texts [8].
We performed a polarity-based sentiment analysis of the previous Long-COVID PASC dataset by using the VADER TOOL. First, each text was translated in English using the API google (Python library), then the VADER tool[3] was applied to each of the 73 texts and a polarity value (ranking between -1,0,1) was associated to each of them. A lexicon-based approach for extracting polarity from written texts in the English language is the VADER tool, namely the Valence Aware Dictionary and sEntiment Reasoner [7]. VADER combines highly curated lexicon resources with rule-based modeling consisting of 5 humans validated rules to predict a polarity score from textual input.
Several are the advantages of a VADER-like approach for clinical applications. One of the most important is that VADER does not require a training phase. Consequently, it can be easily applied in both low-resource domains and in fields where a limited number of data can be collected at a time, as usually happens in clinical trials.
The original VADER has been proven to work well on short text, making it usable to analyze open answers to questionnaires, which still are the most widespread method for carrying out investigations in the clinical field.
Moreover, a lexicon-based approach is generally faster than other approaches and, therefore, it may be suited for near real time applications. In addition, by exploiting general "parsimonious" rules, a VADER-like approach is domain-agnostic and easily customizable to different domains by extending the existing dictionary with domain-specific words. Finally, it can be considered a white box model, which makes it highly interpretable and highly adaptable to different languages.
The starting point of the VADER system is a generalizable, valence-based, human-curated gold standard sentiment lexicon, built on top of three well-established lexicons, i.e. LIWC [17], General Inquirer and ANEW [3], expanded with a set of lexical features commonly used in social media, which included emoji, for a total of 9,000 English terms subsequently annotated in a [-4, 4] polarity range through the Amazon Mechanical Turk's crowd-sourcing service. The second core step of the VADER engine is the identification of some general grammatical and syntactical heuristics to identify semantic shifters.

[3] https://github.com/cjhutto/vaderSentiment

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65

3.3 Topic Modeling using LDA

We performed a topic model analysis of the previous Long-COVID PASC dataset by iterative appling LDA[4]. A topic can be seen as a collection of representative words in a text, that helps to identify what is the subject or the subjects of a document. Latent Dirichlet Allocation (LDA) [5] is a generative probabilistic model commonly used for the identification of latent topic in textual corpora. The starting point of an LDA model is that a document is represented as a bag-of-words. Each single word is represented as a pair of values: the first represents its position, the second is the number of occurrences of the word itself within the document. The assumption under an LDA model is that each document in a corpus can be modeled as a mixture of a finite number of topics with a certain probability, while a topic can be characterized by a distribution over words. More in details, assume that there are k topics across all documents. Le $\mathbf{w}$ be a document in a corpus $\mathcal{D}$ and let consider:

- $\theta \sim Dir(\alpha)$, be a mixture of k topics, having a Dirichlet probability distribution where α is the per-document topic distribution;
- a topic $z_n \sim Multinomial(\theta)$, where n represents the number of words that define a topic;

then, given the topic z_n, a word w_n is sampled from $p(w_n|z_n, \beta)$, and β represents the per-topic word distribution. Then, the probability of $\mathbf{w}$ containing n words can be described as (Equation 1):

$$p(\mathbf{w}) = \int \left(\prod_{n=1}^{N} \sum_{z_n=1}^{k} p(w_n|z_n, \beta) \right) p(\theta, \alpha) d\theta \tag{1}$$

4 Results and discussion

To gain useful insights about the collected data, a preliminary exploratory data analysis was performed by summarizing data through suitable visualization.
In this section, we present the results of three different approaches applied on the COVID-19 narrative dataset. First part of the analysis corresponds to the extraction of the polarity using the VADER software, in which, starting from the score, the polarity was inferred by dichotomizing the overall score. In the second part, the topic modeling was applied through LDA, extracting a number of topics equal to three. The third phase is based on the objective of finding a qualitative relationship between the polarity extracted with VADER and the latent topics with LDA, considering this as a semi-supervised problem. Once polarity was considered as a real label of the various texts, the presence of polarized topics was explored and therefore a correspondence between the sentiment and the topic was found.

[4] https://github.com/topics/lda-topic-modeling

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65

4.1 Exploratory data analysis

The longest document among PASC narrative counts 660 words and an interquartile range of 366.5 words [24]. The most used words in the documents are shown through a word cloud, as shown in Fig. 2.

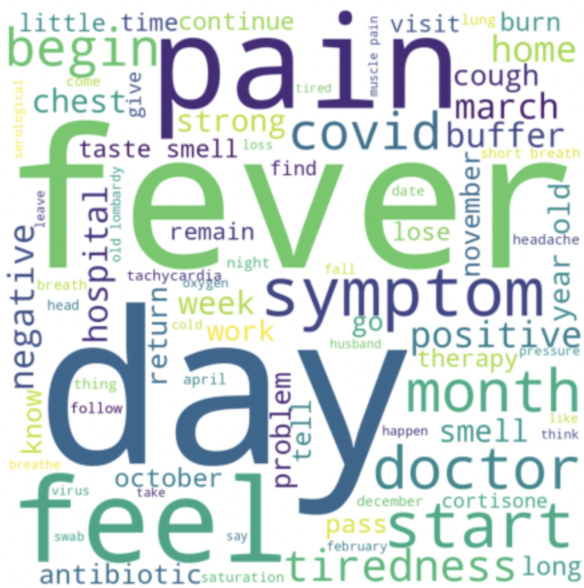


Fig. 2. Word cloud showing the most frequent tokens in the document after preprocessing. Tokens with the largest font size are the most frequent.

The corpus was pre-processed with standard NLP techniques, i.e., tokenization, stop word removal, and lemmatization and then the word cloud was built by a Python word cloud generator. Terms as "fever", "doctor", "pain", "covid", "day" are the most frequent words in the corpus, followed by "tiredness", "symptom", "doctor", "antibiotic", "positive", "negative".

4.2 Polarity analysis

Initially the dataset did not have a polarity score. VADER has been used on the dataset containing 73 narrations to extract the polarity by associating a score equal to 1 for positive polarity and -1 for negative polarity. The results show that, by calculating the occurrence of positive and negative polarities, VADER assigned positive polarity (POS) to 12 documents and negative polarity (NEG) to 61 documents, as demonstrated in Table 1. The application of VADER revealed more negative polarities than positive ones, as shown in Fig. 3.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65

Table 1. The total number of documents is 73, 12 tagged as positive and 61 as negative.

N.o of Documents	POS	NEG
73	12	61

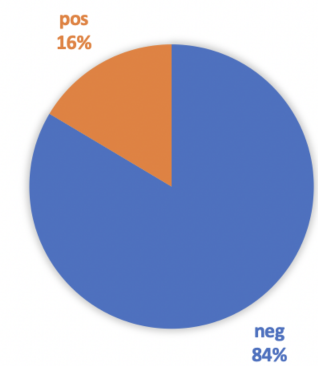


Fig. 3. Pie chart of polarities. The 73 documents were labeled as positive or negative using VADER.

4.3 Topic modeling analysis

In this task, topic modeling was applied through LDA, extracting a number of topics equals to 3. Each topic is a combination of 10 keywords, and each keyword contributes a certain weightage to the topic. Table 2 contains for each topic the keywords with the associated weight.

To continue the analysis, the number of documents associated with the main topic to which they belong was obtained using the python library *gensim*, as shown in Table 3.

To verify the presence of polarized topics, the 12 documents with positive polarities were selected (see Fig. 3), obtaining the assignment of the polarities to the specific topic, as shown in Table 4.

Then, the analysis of polarized topics was carried out again with reference to the 61 PASC testimonials with negative polarities, as shown in Table 5.

The results show that the 7 positive documents are associated with topic 0, 4 are associated with topic 1 and the remaining document with topic 2.

The results show that the split is 1/3 per topic. In fact 20 negative documents are associated with topic 0, other 20 with topic 1 and the remaining 21 documents with topic 2.

Finally, in the last stage, the keywords of each topic were entered into a list to be analyzed with VADER. Each was associated not only with positive negative polarity scores, but also with neutral ones. Note that although there is a distinction in positive and negative labels regarding the texts, the topics have in descending order predominantly a neutral, negative polarity, while positive equal to 0 for each topic, as shown in Table 6.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65

Table 2. Keywords per topic with weights.

Topic ID	Keywords and weights
Topic 0	0.019 · fever + 0.014 · doctor + 0.030 · pain + 0.012 · year + 0.011 · home + 0.011 · hospital + 0.011 · start + 0.010 · symptom + 0.009 · begin + 0.009 · march
Topic 1	0.029 · pain + 0.013 · start + 0.013 · year + 0.012 * fever + 0.010 · negative + 0.010 · chest + 0.009 · covid + 0.009 · feel + 0.009 · smell + 0.009 · month
Topic 2	0.022 · pain + 0.017 · smell + 0.016 · feel + 0.015 · year + 0.015 · fever + 0.015 · covid + 0.013 · taste + 0.011 · negative + 0.011 · month + 0.010 · symptom

Table 3. The total number of texts for the associated topic.

Topic	N. Texts	%
0	27	37%
1	24	33%
2	22	30%

Table 4. Documents with positive polarity, partitioned with respect to the 3 topics.

Topic	Positive N. Texts	%
0	7	59%
1	4	33%
2	1	8%

Table 5. Documents with negative polarity, partitioned with respect to the 3 topics.

Topic	Negative N. Texts	%
0	20	33%
1	21	34%
2	21	34%

Table 6. Polarity score and class polarity for each topic. Polarity scores were extracted with VADER, while polarity classes are the result of a discretization process which associates the positive polarity (POS) for scores greater than 0 and negative polarity (NEG) for scores less than 1.

Topic	POLARITY SCORE	CLASS POLARITY
0	-0.510	NEG
1	-0.79	NEG
2	-0.79	NEG

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65

5 Conclusions and Future Work

In the present work, some sentiment analysis and topic modeling methods have been applied to 73 Italian narrative medicine texts of patients with post-acute sequelae of COVID-19, i.e. PASC. After carrying out the automatic collection and translation in English of the texts, VADER software was applied in order to extract the polarity of the 73 narratives, resulting in 61 negative and 12 positive documents. The results of this approach seem more consistent with the theme of Long-COVID, as it seems intuitively more correct that the testimonies have a predominantly negative polarity.
With the aim of finding a connection between the extracted polarities and possible topics contained in texts, topic modeling was applied to extract the latent topics through the LDA model. In order to conduct a qualitative analysis and find a relationship between the polarity and the related topics, it was checked whether the extracted topics mainly mention negative or positive words.
However, the results did not show significant differences between what emerges from the polarity analysis and from the topic extraction, since in both approaches the prevailing polarity is the negative one. It can be concluded that the analysis made at the document level is not exactly representative of the true feeling of the texts.
Better results may be obtained by analyzing the texts with a sentiment analysis approach not of the "document level" type, but of the "sentence/aspect level" type. In future work we plan to extend the dataset and consider the sentence level approach to perform sentiment analysis.
An other interesting future development would be to apply Named Entity Recognition analysis on the proposed dataset by finding a possible correlation with the polarity analysis.

References

1. Agrusta, M., Cenci, C.: Telemedicine and digital narrative medicine for the customization of the diagnostic-therapeutic path at the time of covid 19 (2021)
2. Boon-Itt, S., Skunkan, Y.: Public perception of the covid-19 pandemic on twitter: Sentiment analysis and topic modeling study. JMIR Public Health and Surveillance **6**(4), e21978 (2020)
3. Bradley, M.M., Lang, P.J.: Affective norms for english words (anew): Instruction manual and affective ratings. Tech. rep., Technical report C-1, the center for research in psychophysiology ... (1999)
4. Chandrasekaran, R., Mehta, V., Valkunde, T., Moustakas, E.: Topics, trends, and sentiments of tweets about the covid-19 pandemic: Temporal infoveillance study. Journal of medical Internet research **22**(10), e22624 (2020)
5. Hidayatullah, A.F., Aditya, S.K., Gardini, S.T., et al.: Topic modeling of weather and climate condition on twitter using latent dirichlet allocation (lda). In: IOP Conference Series: Materials Science and Engineering. vol. 482, p. 012033. IOP Publishing (2019)
6. Hossain, M.M., Tasnim, S., Sultana, A., Faizah, F., Mazumder, H., Zou, L., McKyer, E.L.J., Ahmed, H.U., Ma, P.: Epidemiology of mental health problems in covid-19: a review. F1000Research **9** (2020)

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65

7. Hutto, C., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the international AAAI conference on web and social media. vol. 8, pp. 216–225 (2014)
8. Jalil, Z., Abbasi, A., Javed, A.R., Badruddin Khan, M., Abul Hasanat, M.H., Malik, K.M., Saudagar, A.K.J.: Covid-19 related sentiment analysis using state-of-the-art machine learning and deep learning techniques. Frontiers in Public Health **9**, 2276 (2022)
9. Kim, L., Fast, S.M., Markuzon, N.: Incorporating media data into a model of infectious disease transmission. PloS one **14**(2), e0197646 (2019)
10. Kiritchenko, S., Zhu, X., Mohammad, S.M.: Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research **50**, 723–762 (2014)
11. Maison, D., Jaworska, D., Adamczyk, D., Affeltowicz, D.: The challenges arising from the covid-19 pandemic and the way people deal with them. a qualitative longitudinal study. PloS one **16**(10), e0258133 (2021)
12. Mansoor, M., Gurumurthy, K., Prasad, V., et al.: Global sentiment analysis of covid-19 tweets over time. arXiv preprint arXiv:2010.14234 (2020)
13. Martinis, M.C., Zucco, C., Cannataro, M.: An italian lexicon-based sentiment analysis approach for medical applications. In: Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. pp. 1–4 (2022)
14. Medford, R.J., Saleh, S.N., Sumarsono, A., Perl, T.M., Lehmann, C.U.: An "infodemic": leveraging high-volume twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. In: Open forum infectious diseases. vol. 7, p. ofaa258. Oxford University Press US (2020)
15. Mehandru, S., Merad, M.: Pathological sequelae of long-haul covid. Nature immunology **23**(2), 194–202 (2022)
16. Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., Gao, J.: Deep learning–based text classification: a comprehensive review. ACM computing surveys (CSUR) **54**(3), 1–40 (2021)
17. Pennebaker, J.W., Boyd, R.L., Jordan, K., Blackburn, K.: The development and psychometric properties of liwc2015. Tech. rep. (2015)
18. Pye, A., Roberts, S.R., Blennerhassett, A., Iqbal, H., Beenstock, J., Iqbal, Z.: A public health approach to estimating the need for long covid services. Journal of Public Health (2021)
19. Qorib, M., Oladunni, T., Denis, M., Ososanya, E., Cotae, P.: Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on covid-19 vaccination twitter dataset. Expert Systems with Applications **212**, 118715 (2023)
20. Rosenberg, H., Syed, S., Rezaie, S.: The twitter pandemic: The critical role of twitter in the dissemination of medical information and misinformation during the covid-19 pandemic. Canadian journal of emergency medicine **22**(4), 418–421 (2020)
21. Rossi, R., Socci, V., Talevi, D., Mensi, S., Niolu, C., Pacitti, F., Di Marco, A., Rossi, A., Siracusano, A., Di Lorenzo, G.: Covid-19 pandemic and lockdown measures impact on mental health among the general population in italy. Frontiers in psychiatry p. 790 (2020)
22. Rushforth, A., Ladds, E., Wieringa, S., Taylor, S., Husain, L., Greenhalgh, T.: Long covid–the illness narratives. Social Science & Medicine **286**, 114326 (2021)
23. Scarpino, I., Zucco, C., Cannataro, M.: Characterization of long covid using text mining on narrative medicine texts. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 2022–2027. IEEE (2021)

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65

24. Scarpino, I., Zucco, C., Vallelunga, R., Luzza, F., Cannataro, M.: Investigating topic modeling techniques to extract meaningful insights in italian long covid narration. BioTech **11**(3), 41 (2022)
25. Sun, C., Yang, W., Arino, J., Khan, K.: Effect of media-induced social distancing on disease transmission in a two patch setting. Mathematical biosciences **230**(2), 87–95 (2011)
26. Taquet, M., Luciano, S., Geddes, J.R., Harrison, P.J.: Bidirectional associations between covid-19 and psychiatric disorder: retrospective cohort studies of 62 354 covid-19 cases in the usa. The Lancet Psychiatry **8**(2), 130–140 (2021)
27. Tchuenche, J.M., Bauch, C.T.: Dynamics of an infectious disease where media coverage influences transmission. International Scholarly Research Notices **2012** (2012)
28. Umair, A., Masciari, E., Habib Ullah, M.H.: Sentimental analysis applications and approaches during covid-19: a survey. In: Proceedings of the 25th International Database Engineering & Applications Symposium. pp. 304–308 (2021)
29. Valdez, D., Ten Thij, M., Bathina, K., Rutter, L.A., Bollen, J.: Social media insights into us mental health during the covid-19 pandemic: longitudinal analysis of twitter data. Journal of medical Internet research **22**(12), e21418 (2020)
30. Wicke, P., Bolognesi, M.M.: Covid-19 discourse on twitter: How the topics, sentiments, subjectivity, and figurative frames changed over time. Frontiers in Communication **6**, 45 (2021)
31. Yeasmin, N., Mahbub, N.I., Baowaly, M.K., Singh, B.C., Alom, Z., Aung, Z., Azim, M.A.: Analysis and prediction of user sentiment on covid-19 pandemic using tweets. Big Data and Cognitive Computing **6**(2), 65 (2022)
32. Zucco, C., Calabrese, B., Agapito, G., Guzzi, P.H., Cannataro, M.: Sentiment analysis for mining texts and social networks data: Methods and tools. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **10**(1), e1333 (2020)

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_65