Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240821

Greek Tweets on Long COVID: Topic Modelling Following Sentiment Analysis and ChatGPT Interpretation

Afroditi KATIKA[a,1] , Emmanouil ZOULIAS[a], Vassiliki KOUFI[b] and Flora MALAMATENIOU[a]
[a] *Department of Nursing, National and Kapodistrian University of Athens, Athens Greece*
[b] *Department of Digital Systems, School of Information and Communication Technologies, University of Piraeus, Piraeus Greece*

Abstract. Long COVID is a disease that makes it hard for patients to get an official diagnosis while it impacts their quality of life. Many people are turning to social networks such as Facebook, WhatsApp, Twitter (now X) to express their opinions and feelings regarding Long COVID. In this paper, positive (or neutral) and negative text messages in the Greek language, posted on the Twitter platform in 2022, regarding Long COVID are analyzed and popular discussion topics are extracted. Analysis revealed that when topic modelling follows sentiment analysis more coherent topics are created. Furthermore, ChatGPT is used to assign a label to each topic that, in turn, is assessed by a human expert.

Keywords. Big Data, Twitter, Natural Language Processing, LLM, COVID-19

1. Introduction

Long COVID is a syndrome with a diverse set of symptoms. Patients often do not show residual disease or cardiopulmonary complications making it hard to get treated or get an official diagnosis [1]. Hence, patients often are turning to social media and platforms like Twitter (now X) to express their opinions and emotions and look for other patients in similar situations. Natural Language Processing (NLP) techniques are being employed to monitor what people are saying. Specifically, Topic Modelling methods have been extensively used to reveal popular discussion topics in social media especially in the advent of COVID-19 [2].This paper presents an approach leveraging the Large Language Model ChatGPT. The most relevant discussed topics regarding Greek tweets on long COVID, per positive and negative sentiment, are extracted and interpreted using OpenAI's ChatGPT [3].

[1] Corresponding Author: Afroditi KATIKA, Athens, Greece; E-mail: afro.katika@gmail.com.

2. Methods

The dataset included 930 tweets that were posted between January 1st and 31st December 2022 on Twitter (now X), in Greek language. Tweets were classified upon their sentiment in positive/neutral and negative using Greek-BERT[4]. After creating one separate dataset for each sentiment, tweets were pre-processed and lemmatized. Latent Dirichlet Allocations (LDA) technique was used for topic modelling. Gensim's open-source library was employed to reveal discussion topics per sentiment. After tokenising tweets, a dictionary was created. In order to interpret the topics, Open AI ChatGPT 3.5 was used [3]. The labels were rated by a post-graduate student in the area of Healthcare Informatics as: a) Useful, b) Moderately Useful, c) Not really useful, and d) Not useful.

3. Results and Discussion

Coherence score varied between 0.50-0.59 for positive or neutral topics and 0.42-0.56 for negative topics. The results of the human assessment of the ChatGPT labels: 40% of the labels were considered to be Useful, 40% were considered Moderately useful and 20% Not useful. Compared to a previous study [4], higher coherence scores show that Topic Modelling performs better when the tweets share the same sentiment. 80% of all ChatGPT labels were deemed useful by the expert. Compared to an earlier study they are higher [5] potentially due to the use of ChatGPT.

4. Conclusions

The analysis indicates that public health authorities could create an end-to-end pipeline to mine knowledge from social media and quickly understand main discussion topics of a new disease. Extracting topics after grouping tweet messages per sentiment would make next steps more targeted. Further studies in this area could investigate a bigger range of emotions as well as use Natural Language Processing techniques that perform more accurately in the Greek language.

References

[1] Manta A, Michelakis I, Dafni M, et al. Long-term outcomes, residual symptoms and quality of life in COVID-19 hospitalized patients: A 12-month longitudinal study. *https://doi.org/101177/10815589231212899*. Epub ahead of print 29 November 2023. DOI: 10.1177/10815589231212899.

[2] Tsao SF, Chen H, Tisseverasinghe T, et al. What social media told us in the time of COVID-19: a scoping review. *Lancet Digit Heal* 2021; 3: e175–e194.

[3] ChatGPT, https://openai.com/chatgpt (accessed 3 March 2024).

[4] Katika A, Zoulias E, Koufi V, et al. Mining Greek Tweets on Long COVID Using Sentiment Analysis and Topic Modeling. *Stud Health Technol Inform* 2023; 305: 545–548.

[5] Scheepers F, Zervanou K, Spruit M, et al. Towards Interpreting Topic Models with ChatGPT. Paper presented at The 20th World Congress of the International Fuzzy Systems Association, www.tue.nl/taverne (accessed 6 September 2023).