CyberPharmacovigilance of Covid-19: Social Media Data Analytics

Eustache Muteba AYUMBA[a,1]
aInternational Medical Informatics Association, Correspondent in DR Congo

Abstract. In this article, we present a methodology and a tool for extracting and analyzing information, reported by a social media monitor, of people who have taken drugs to treat Covid-19 and the adverse effects encountered.

Keywords. Pharmacovigilance, Covid-19, Social media monitoring, Data analytics

1. Introduction

WHO's involvement in monitoring the situation of pharmaceutical industry is pushing the authorities in charge of the issue of organizing pharmacovigilance. Most of time, decisions in the area of drug safety are based on available information from pharmacovigilance systems, which are mostly based on Adverse Drug Reactions (ADR) reports, voluntarily made by healthcare professionals and consumers. In some cases, some decision-makers and health caregivers do not have up-to-date and well-organized pharmacovigilance systems. With the onset of the COVID-19 pandemic, social media has rapidly become a crucial communication tool for information generation, dissemination, and consumption. Thus, social media monitoring [1], along with conventional pharmacovigilance measures, can be used to detect signals associated with any information related to adverse effects or any other problem related to marketed medicines. This article presents a methodology and a tool for extracting and analyzing information, reported by a social media monitor, of people who have taken drugs to treat Covid-19 and the adverse effects encountered.

2. Methods

Our analyzer system is naturally based on existing works such as developed in [2, 3, 4 and 5]. It is composed by the following components: the social media monitor, the social media report file, the indexer, the search index, the index file and the dictionary, the user interface, the extractor engine, the pharmacovigilance database and the data analytics report . The Analyzer system tasks are the following. Build the indexer (index file) from the lexicons and set a dictionary; the user interface allows the user to submit queries; once the query is processed, the extractor engine sends the results to the user interface for validation; on needed refines information; stores it in the database and reports the data analytics result. The extraction and the analysis of data are based on a query text as

---

[1] Eustache Muteba Ayumba, P.O. Box 14769, 01, Kinshasa, Congo, E-mail: emuteba@hotmail.fr

decision criteria or rules. The extractor engine captures salient information and computes the term frequency (TF) of the indicators (keywords) retained in the analysis plan. We proposed the following query structure. *Query Structure (Expression form):*

(1) Categories: keywords and basic units that identify the conversation
(2) Lexicons: keywords that specify your query with semantic type avoiding noisy.
(3) Boolean terms: "AND, OR, and NOT" that create more powerful searches.

3. Results and Discussion

The social media monitoring tool used in our study is Alerti (https://fr.alerti.com/). Our analysis plan was based on query text containing keywords such as categories are "anti-covid", "side effect", "adverse drug reaction" and others keywords related to drug used as anti-covid. The social media monitor generates a report in .csv file. Our analyzer system, pharmavigil (https://www.vchf.net/pharmavigil/pharmav.php), was design and build to reinforce the data analysis, with the same query text, to get salient information. The experimentation was done on the social media monitor report (file). Our data analytics report generated by the extractor engine provides two kinds of information namely a list of record lines containing keywords and the term frequencies. We note, in case of anti-covid drug "ivermectin", the TF value is 1390 on a total 1756 of record lines. And a "nitazoxanide" drug has a TF value 0 on the same total records lines. The task of analyzing data and searching specific words into a social media monitoring report is subject to many challenges. One of the crucial challenges is a language style devoid of linguistic vigor. It should be noted that to overcome the limits imposed by the keyword search, our system uses the dictionary. This allows so little to eliminate the noises therefore the post messages containing irrelevant information.

4. Conclusion

On the strength of the above, we believe that in the field of health, social media can be a potentially exploitable source of information to complete and optimize pharmacovigilance processes. The main contribution of our study is the methodology of developing an analytics supplement tool that aggregate data coming from social media monitoring.

References

[1]   Lee J, Lee Y, Kim DH, Lee HS, Yang BR, Kim MG. The Use of Social Media in Detecting Drug Safety–Related New Black Box Warnings, Labeling Changes, or Withdrawals: Scoping Review. JMIR Public Health Surveill. 2021 Jun;7(6):30137.

[2]   Twetman H, Paramonova M, Hanley M. Social Media Monitoring: A Primer - Methods, tools, and applications for monitoring the social media space, NATO Strategic Communications Centre of Excellence. 2020.

[3]   Podhoranyi, M. A comprehensive social media data processing and analytics architecture by using big data platforms: a case study of twitter flood-risk messages. Earth Sci Inform. 2021; 14:913–929.

[4]   Mucchielli R. L'Analyse de Contenu - Des Documents et des Communications. 5ième ed. Paris : ESF; 1984.

[5]   Levene M. An Introduction to Search Engines and Web Navigation. 2nd ed. New Jersey: John Wiley & Sons; 2010.