Science Recommendations Based on Social Media Platforms

Jorge Oliveira e Sá
Researcher, ALGORITMI Research Center
University of Minho
Guimarães, Portugal
jos@dsi.uminho.pt

Abstract

Social media platforms can be a source of recommendations for various items or products. Digital Marketing takes advantage of this potential, especially when using influencers to evaluate their items or products on social media platforms. Each influencer can be a recommender of items or products. Thus, one can question why this does not occur similarly in the area of science. This work presents a study in which shares/mentions were collected on a social media platform, Twitter, about scientific documents (for example, journal or conference papers, thesis, reports, and patents). We pretend to obtain the polarity of these shares/mentions (positive, negative, and neutral) is obtained. It was found that most of the shares/mentions are neutral and few are considered positive, which leads to the conclusion that there are shares/mentions about science and that, despite the lack of detected polarity they can serve as a basis for a scientific recommendation.

Keywords
Science Recommendation System; Social Media Platform; Twitter; Sentiment Analysis; Science.

1. Introduction
The sharing of scientific knowledge is traditionally carried out through publications in scientific journals or books, subject to review, editing, printing, and distribution with long waiting periods between each step, access restrictions, and very high membership costs. The development and proliferation of the Internet as an open, free and global communication platform presents great potential for science communication. The Internet revolutionized access to science and made it possible to bridge the geographical barriers between researchers from different countries and research centers, contributing to the development of more effective and productive communication, and guaranteeing collective science (Montgomery, 2009). The development of science itself was thus changed with the integration of the scientific community in a global network (Silva, 2002) standardized by fields of activity and the constitution of transdisciplinary networks of collaboration between different scientific entities and researchers from different areas and countries, being reality consortia and partnerships to carry out projects at European and/or international level. The flow of scientific communication thus includes the production and publication of the results of the research carried out, dissemination to the scientific community and the general public; access to published literature, informal communication between researchers themselves, communication with the public, and incorporation of new knowledge (Morris & Ogan, 1996). Watermeyer (2010) and Bik & Goldstein (2013) refer that online contact via social networks is a two-way contact and allows an informal connection between scientists and between them and the public.

Science can thus be enriched through different representations of scientific knowledge, a multiplicity of opinions, and criticisms that are constructed and shared throughout the network. In short, the phenomenon of social networks, as a natural consequence of the social character of Man, stands out in the technological means of communication and social interaction as a potential tool for accessing, sharing, and disseminating knowledge.

In this way, the increase in scientific publications (in this paper, scientific publications will be called research outputs (RO), thus, an RO can be a journal paper, a conference paper, a book, a chapter of a book, a scientific report, a patent, among others) in recent years has led to the need for recommendations on which ROs to read. Usually, the ROs go through peer-review processes. The result of the peer-review work is usually not made public, creating an opacity in the perception of the relevance of the ROs. Finding the right RO for those looking for a high number of published ROs is a complex and arduous task, especially if the person/researcher looking for it is still a beginner in a certain area of knowledge. The use of social media platforms means that bibliometric indicators, namely the number of citations, traditionally used to assess the relevance of a RO, are complemented with indicators called altmetrics (alternative metrics) based on: shares on social media, likes, and views. Altmetrics allows to measure the attention that an RO is causing both in society in general, and in the scientific community (this was observed in this time of pandemic), in this way, they can help researchers in their research processes to identify the most relevant ROs.

1.1 Objectives

This paper aims to collect shares and mentions made on the social media platform, Twitter, to understand if tweets are used as a means of disseminating science and especially if the scientific community uses social media platforms (in this case Twitter) to mention ROs. For these mentions, their polarity will be analyzed to identify the positive, neutral, and negative ones. Thus, it will be possible to verify whether these mentions can serve as the basis of a RO recommendation system supported by the researchers' experience and their social involvement in a scientific community.

2. Literature Review

2.1. Recommendation

The Internet and web services have increased exponentially in recent decades; a surplus of information is now accessible to everyone. It can be a challenge for users to filter all this information to extract its essential aspects. Many online e-commerce companies recommend products to their users, selling millions of products through a web platform. For an average user, navigating through all these possibilities can be exhausting work. Recommender systems aim to solve this information overload problem while personalizing the user experience by delivering accurate and personalized recommendations of items/products according to their preferences.

Opinion is fundamental in almost all human activities because it is through them that people's behavior is influenced. Customers want to know the opinion of others about a particular product before they buy it and they use social media platforms to obtain these opinions (Keith, Lettura, & Villegas, 2019).

2.2. Recommendation Algorithms

Collaborative filtering evaluates items/products using ratings (implicit or explicit) given by users. If a user A and a user B rate items/products in common, then their preferences are considered similar. Therefore, if there are items/products in user A's history that are not in user B's history, then these items/products can be recommended to user B. This, collaborative filtering is the process of recommending items/products using the opinions of others (Xu, Jiang, Chen, Ren, & Liu, 2019). There are two categories of collaborative filtering (Sen & Udgirikar, 2013):

- user-based - the recommendation system uses the profiles of similar users to make the recommendation, they look for nearest neighbors and, according to their interests, the user's interests are obtained (Xu et al., 2019); and
- item-based - if a user positively evaluates an item/product, the recommendation system identifies similar candidate items/products for a posterior recommendation.

Content-based recommendations allow building a user's profile to predict products not seen by the user, using tags and keywords. It does not require data from other users, it only requires the analysis of the items/products and the user profile to make recommendations (Fayyaz, Ebrahimian, Nawara, Ibrahim, & Kashef, 2020).
The generation of a recommendation starts with the selection of the N most relevant items/products for the user and then looks for the similarity between the existing characteristics in the user profile and the items/products (Jomsri, Sanguansintukul, & Choochaiwattana, 2010). Other works present changes in this algorithm, as, in addition to the most relevant items/products, they provide fortuitous recommendations (serendipity) that are items/products from different or distant areas (Sugiyama & Kan, 2011), but which can be useful to users. The last step of the recommendation system is to sort the elements of this list in a certain order and the N elements at the top of this list will be recommended to the user.

There are also other recommendation algorithms, namely:

- based on demographics – allows recommending items/products based on a user's demographic characteristics: age, gender, nationality, occupation, etc. Generally, they collect this data during user registration, where a series of items/products are also presented to be evaluated to build their profile (Al-Shamri, 2016);
- utility-based – allows recommending items/products with greater utility for the user, based on the calculated utility, based on the Multi-Attribute Utility Theory (Deng, 2015; Fayyaz et al., 2020);

- knowledge-based – allows for recommended items/products based on the user's preferences, these systems do not need much information about a user since they are independent of the user's tastes and are also independent of the user's evaluations, in this way they use the explicit knowledge of the items/products and users to generate recommendations (Devchand, Sheehan, Gallivan, Tuncer, & Nicols, 2017). These systems are seen as independent systems and can also be considered complementary to other types of recommender systems (Fayyaz et al., 2020);
- hybrids – a combination of two or more previous techniques to obtain better performance and accuracy in the recommendation. Its main objective is to eliminate the individual disadvantages of each of the techniques presented (Fayyaz et al., 2020). For example, a collaboration-based recommendation with a content-based recommendation (Fayyaz et al., 2020), allows the recommendation to decrease sensitivity to the number of users who rated the item (Sarwar, Karypis, Konstan, & Riedl, 2002).

2.3. Mention-based recommendation system
This recommendation system stands out, as it is based on ratings and mentions of items/products shared by users on social platforms. These systems are used to recommend hotels, restaurants, and other related services (Ramzan, Bajwa, Jamil, & Mirza, 2019). The system works in two parallel ways:

1. the numerical ratings and votes of the items/products are collected and harmonized;
2. the texts of the mentions are processed through Natural Language Processing (NLP) to obtain the polarity. If the polarity value of an item/product characteristic is greater than 0, then that characteristic is evaluated positively, if it is equal to or less than 0, it is evaluated negatively. Thus, the polarity of an item/product mentions can be calculated by summing the polarity of the mentions of all the characteristics of an item/product.

2.4. Recommendation systems based on Influencers
This recommendation system appears with the Internet to reach consumers, as it is much more difficult to reach them through traditional means of communication. Today's consumers have an extremely wide environment and a range of different items/products. This fact stimulates the perception of consumption as an important part of social life. The growth of social networks has completely renewed the way people interact, communicate and engage (Arora, Bansal, Kandpal, Aswani, & Dwivedi, 2019). Social networks have become great platforms for sharing personal information, news, photos, and videos and are essential communication platforms as they facilitate interactions between people, which can influence or affect the opinions of others (Bamakan, Nurgaliev, & Qu, 2019). Research and applied evidence suggest that online opinion leaders are important promoters of products and services in different areas of business and in marketing itself (Byrne, Kearney, & MacEvilly, 2017; Dhanesh & Duthler, 2019). Thus, a more recent concept emerges here: the influencer.

This concept is understood as the digital marketing practice that takes advantage of online users with a good range of followers, who can influence consumer attitudes and decision-making processes in favor of brands or ideas (Wielki, 2020). According to Kartajaya, Kotler, & Hooi (2019) influencers have a large number of followers and their opinions influence these followers. It is also accepted that influencers can create content (Babin & Hulland, 2019) to build their reputation and are considered experts in their communities of followers. Influencers can also be seen as powerful human brands that positively influence the performance of organizations associated with them (Ki, Cuevas, Chong, & Lim, 2020). Organizations have found value and utility in using influencers by making them spokespersons for their brands as they can be seen as opinion leaders. More and more organizations allocate more financial resources to influencer marketing. A study (Nielsen, 2016) indicates that activities involving influencers contribute to increased product sales. Thus, some indicators should be analyzed: 71% of consumers are more likely to make purchases of products based on mentions on social networks (Ewing, 2011); 78% of consumer purchases are impacted by publications on social networks (Olenski, 2012). Another study indicates that 80% of respondents made purchases recommended by influencers through a link or image provided by them (Rakuten, 2019). Thus, an influencer is a recommender, because when people follow an influencer, they feel some kind of trust and pleasure with the influencer content and due to that, people accept the influencer shares/mentions as recommendations.

2.5. Recommendation systems applied to Science
Collaborative filtering is used to recommend Research Outputs (ROs) based on a tag system (Parra and Brusilovsky, 2010). Variants can be identified, in which the similarity between two users is calculated using Pearson's correlation on their ratings of common ROs (Schafer et al. 2007) or through non-binary probabilistic models, to take advantage of the set of user tags (Klampanos et al. 2009).

The content-based recommendation can be applied to the context of science. ROs can be seen as items and researchers as users. Based on the list of citations of a researcher, it is possible to start building their profile and create a list of preferences or interests through keywords that he/she uses in their ROs. The recommendation system can extract the keywords from the title, abstract, and content of the RO to generate a characterization of the RO so that similar ROs can be grouped later. The characteristics of the ROs can be compared with the interests of the researcher and if there are similarities, there is a recommendation (Bai et al., 2019).

A researcher's profile is constructed by its ROs and the set of associated tags (Jomsri et al., 2010). Researchers with little production will be distinguished from those who have more, which allows for distinguishing between junior and senior researchers (Jomsri et al., 2010; Sugiyama and Kan, 2011; Nascimento et al. 2011).

In the context of science recommendations, hybrid approaches are also used. The combination of several approaches to obtain recommendations aims to improve the performance and accuracy of recommendations (Wang et al. 2018).

The use of influencers in the context of science is still not very common. However, in science, it is usual to have a peer-review process to review ROs, as this process is the main mechanism for controlling the relevance of ROs and it is accepted in the scientific communities. Reviewers can be compared to influencers, as they provide suggestions to authors to correct and improve the RO, whether the RO can be published or rejected (Bornmann, 2011).

3. Methods

This work follows the phases of the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, namely, Data Collection, Data Preparation, Modeling, and Evaluation.

For the CRISP-DM Data Collection phase, the social media platform Twitter was selected, after a preliminary analysis of the data, three steps were identified for the Collection, Validation/Transformation, and Storage of the data, see Figure 1.

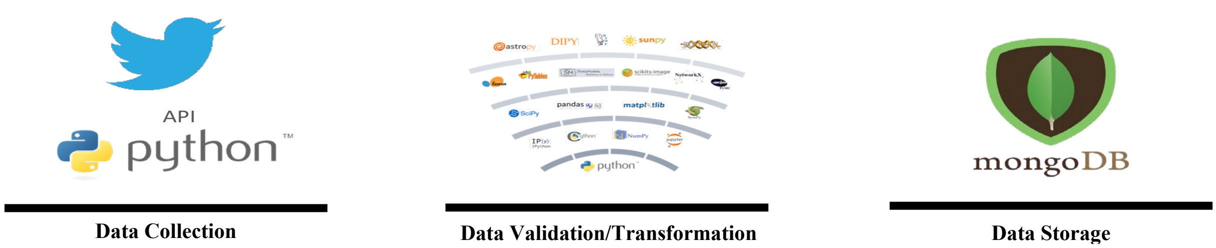

Figure 1. Preliminary steps

Data collection was performed through an Application Programming Interface (API) developed in Python, which uses access credentials provided by Twitter. Access to the API was made through a developer account. After the accesses were configured, the development of the API in Python uses the resources OAuthHandler, StreamListener, and Stream available at Python Tweepy Library. These resources enable to establish a connection and authentication to Twitter, when the authentication is completed, a channel is created to filter and receive the tweets, and the next step is to create a stream that will receive this information. The tweets filtration was made in the fourth step. In a first step, the search was restricted to words such as paper, conference, journal paper, conference paper, thesis, scientific reports, and patents; this search resulted in a set of relevant tweets, but some tweets were unrelated to the source document and needed to be removed. In the second step, it was necessary to study the tweets to find a pattern and it was found that there are tweets that identify where the RO was published and share the link associated with it. The third step was to define keywords that allow identifying publication locations, see Table 1. The fourth and last step results in 30,000 tweets collected.

Table 1. Keywords

Arxiv	ScienceDirect	ProjectMUSE	SSRN	OpenEditionNews
Scopus	Springerlink	UCSBLibrary	EuropePMC_news	WebOfScience
Elsevier	@academia	OCLC	dblp_org	Openlibrary
IEEEXplore	OUPAcademic	ncsulibraries	Ingentaconnect	ElsevierConnect
Plosone	JSTOR	mendeley_com	Nberpubs	NEJM
BioMedCentral,	CambridgeCore	cssci	Directory of Open Access Journals	iopscience
nresearchnews	Pubmed	ibrarycongress	JournalTOCs	TischLibrary
ScienceMagazine	semanticscholar	cinii_jp	PhilPapers	JAMA_current
PNASNews	medlineplus	@APA	ACMDL	AIS

The Data Validation/Transformation started by identifying the original language of the tweets, so as not to discard tweets in other languages, the tweets were translated into English, based on the Python libraries: Langdetect (detects 55 languages) and Googletrans. One of the problems with translating tweets in real time is that some tweets include emojis, causing errors in Googletrans, so it was necessary to exclude emojis from tweets.

Data storage is performed in the NoSQL MongoDB database. Data is stored in real-time through the use of Python's pymongo library. The data was stored in JavaScript Object Notation (JSON).ts.

The CRISP-DM Data Preparation phase consists of preparing the data for sentiment analysis and extraction. In this phase, duplicates are removed, and spam is detected, that is, to detect if a user is commenting on a RO repeatedly to bias the results in a negative or even positive way (untruthful opinions).

To be able to identify the most mentioned ROs or hashtags on Twitter, an algorithm was used that allows each tweet to identify a mention or hashtag, with these mentions and hashtags being placed in lists to identify their occurrences. This algorithm, in addition to identifying mentions and hashtags, also identifies the links of the ROs so that, with the aid of web scraping, it can extract, for example, the title and authors.

Using Python's Natural Language ToolKit (NLTK) library, it is possible to disambiguate tweets by removing punctuation, punctuation introduces noise and adds little value to parsing capability, but removing punctuation from a text makes the unstructured. The tokenization process separates a text into units, such as phrases or words, giving structure to previously unstructured text, for example, the phrase "words in a tweet" is divided into tokens [words, num, tweet]. This task is useful to prepare the text to be treated by a lexical analyzer, which is the next step. After tokenization of the text, we can feed a lexical analyzer to remove stop words, as they are words commonly used in a given language and do not add any value to the data, the NLTK library contains a list of stop words in English. Then the tweet goes through a lemmatization process, that is, it reduces, for example, the words "modeling" and "modeler" to the root word, "model". From the content of the tweets, properly treated with NLP techniques, it is intended to infer whether a tweet presents a positive, negative or neutral opinion about a particular RO. In this way, a sentiment attribution system was defined, which uses the polarity variable from the Textlob library. The polarity variable comprises values between [-1, 1] where -1 defines a negative and 1 a positive sentiment.

The CRISP-DM Modeling phase aims the data interpretation through graphical visualization, making it possible to extract knowledge from the data. The graphical visualizations are explained in section 4 – Work Done.

The CRISP-DM Evaluation phase aims to evaluate and interpret the results obtained in the previous phase (Modeling). To evaluate the results, a presentation was made to a panel of ten researchers, to validate the work and understand the points with a potential improvement. The discussion and results are described in section 5 – Results and Discussion.

4. Work Done

The first visualization consists of the tweet's geographical sources, see Figure 2. As you can see there is a diversity of countries, the collected tweets are mostly sourced in North America, Central America, and Europe. Figure 3 shows the ROs with the most tweets.

Figure 2. Tweets geographical sources

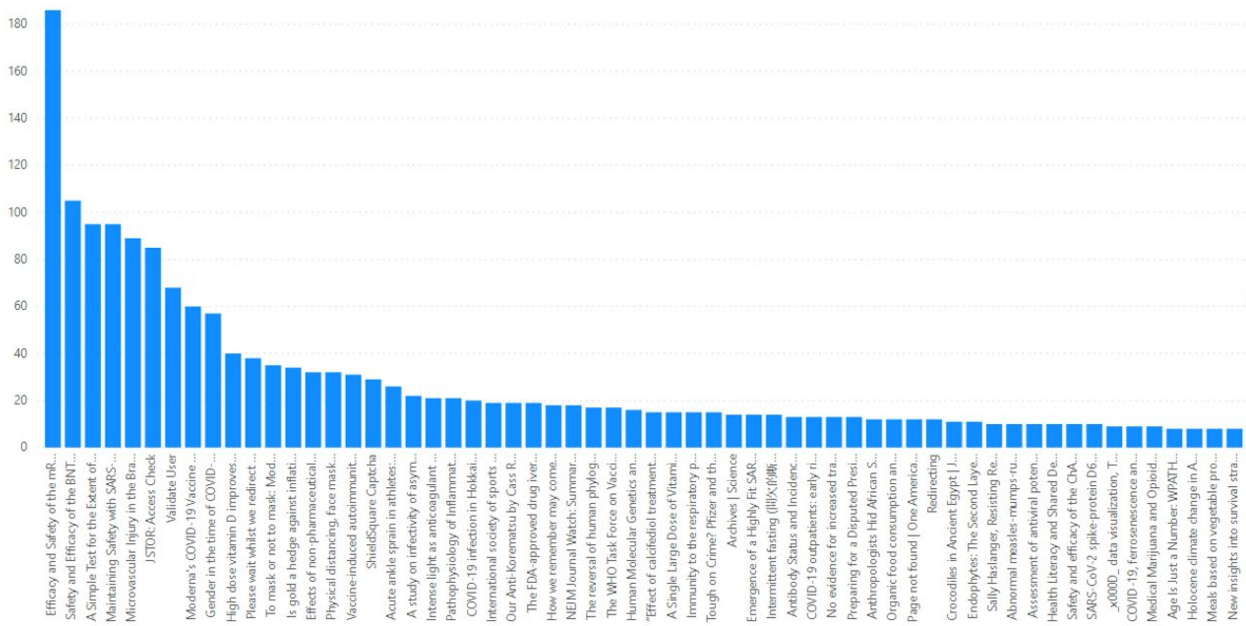

Figure 3. ROs with the most tweets

It can be seen in Figure 3, the most shared RO is "Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine" and in Figure 4 it can be seen that this RO was shared in 15 different languages. It is interesting to verify that a RO can be shared in any geographical place in the world and therefore tweets appear in several languages. Through the analysis of languages, it is possible to verify that English is the predominant language in the tweets, but it is possible to identify that there are tweets in 50 different languages.

One can also visualize the ROs with the most mentions in the tweets and the sentiment associated with these mentions, see Figure 5.

According to the data collected, the most used hashtags, and mentions are related to the pandemic state that is an actual theme, and it is transversal to the whole world, see Figure 6. The places where these ROs are mentioned also confirm this fact, as, for example, the places @JAMA_current, @NEJM, @JAMANetwork, and @sciencemagazine are places where healthcare ROs are shared.

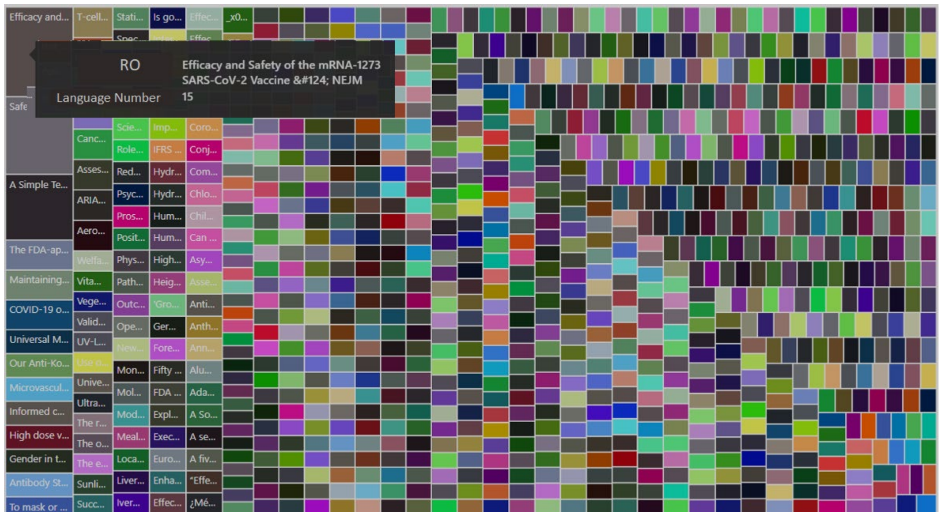

Figure 4. ROs shared in different languages

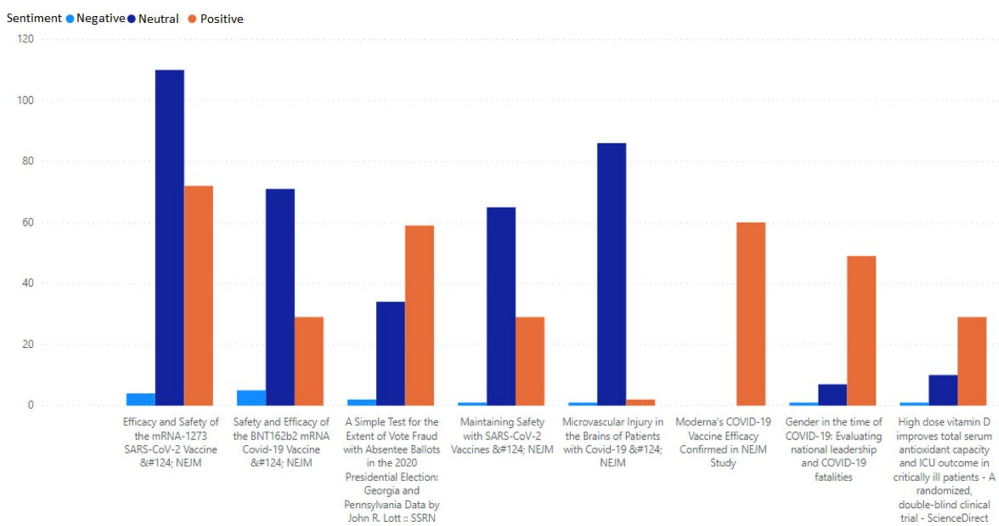

Figure 5. ROs most tweeted and the mention sentiments

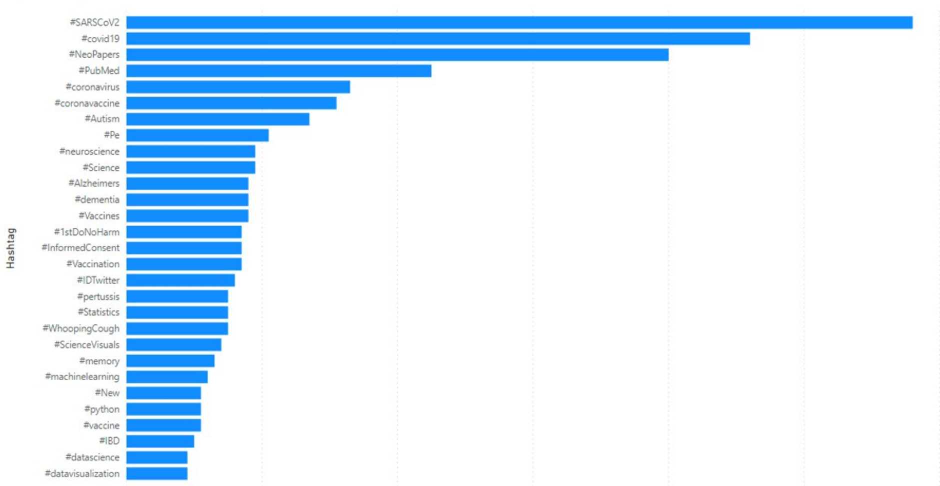

Figure 6. Hashtags and mentions frequency

5. Results and Discussion

As mentioned above, ten researchers were involved to validate the results achieved. Firstly, the researchers questioned the classification of tweets about their sentiment, as there are tweets that are considered positive or negative, but can be considered false positives or negatives.

The perception of the existence of false positives and negatives was performed by reading the text of each tweet and as a result, it was found that the positive tweets were really positive and the negative ones could be considered false negatives, in Table 2 you can read examples of tweets classified as false positives and negatives.

Table 2. Examples of tweets: false positives and false negatives

False Positives	▪ The WHO Task Force on Vaccines for Fertility Regulation. Its formation, objectives and research activities ▪ Meaningful Engagement in the Nursing Home ▪ Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine	NEJM ▪ The Ethical Defensibility of Harm Reduction and Eating Disorders
False Negatives	▪ Health anxiety related to problematic smartphone use and gaming disorder severity during COVID-19: Fear of missing out as a mediator – PubMed ▪ RT @AArieNugraha: Flow-Based Independent Vector Analysis for Blind Source Separation - IEEE Journals & amp; Magazine http://t.co/xWLLe4k42u

To solve this problem, the subjectivity variable was included and considered. This variable allows distinguishing whether the text is a personal or factual opinion. The subjectivity variable is comprised between [0, 1] and quantifies the number of personal opinions and factual information contained in the text, so the value 1 of subjectivity represents that the tweet contains personal opinions and the value 0 indicates that the tweet consists of factual opinions.

For the combination of the variable's polarity and subjectivity, see Figure 7. A tweet to be classified to have a positive sentiment needs the polarity variable to have a value greater than 0.8 and the subjectivity variable to be greater than 0.5. To be sentiment negative, the polarity variable must be less than -0.8 and the subjectivity variable must be greater than 0.5. The remaining values are considered sentiment neutral. This enables to identify the ROs that have more positive, negative, or neutral sentiment evaluations.

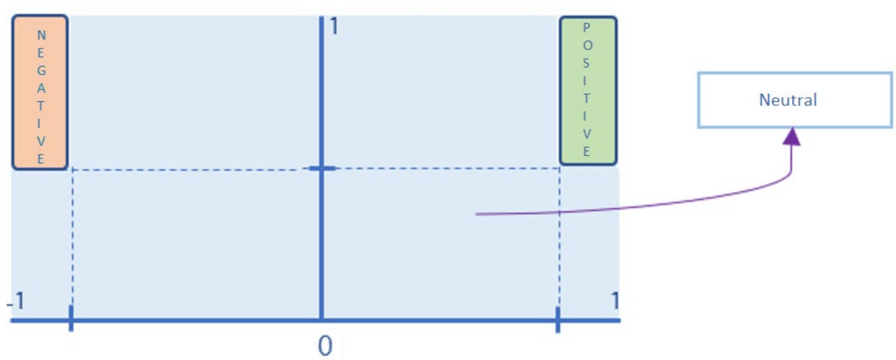

Figure 7. Polarity and subjectivity variables combination

Thus, it appears that considering the subjectivity variable in the sentiment analysis causes a change in the results. It can be seen that the number of negative and positive comments drops considerably, increasing the number of comments considered neutral, see Figure 8.

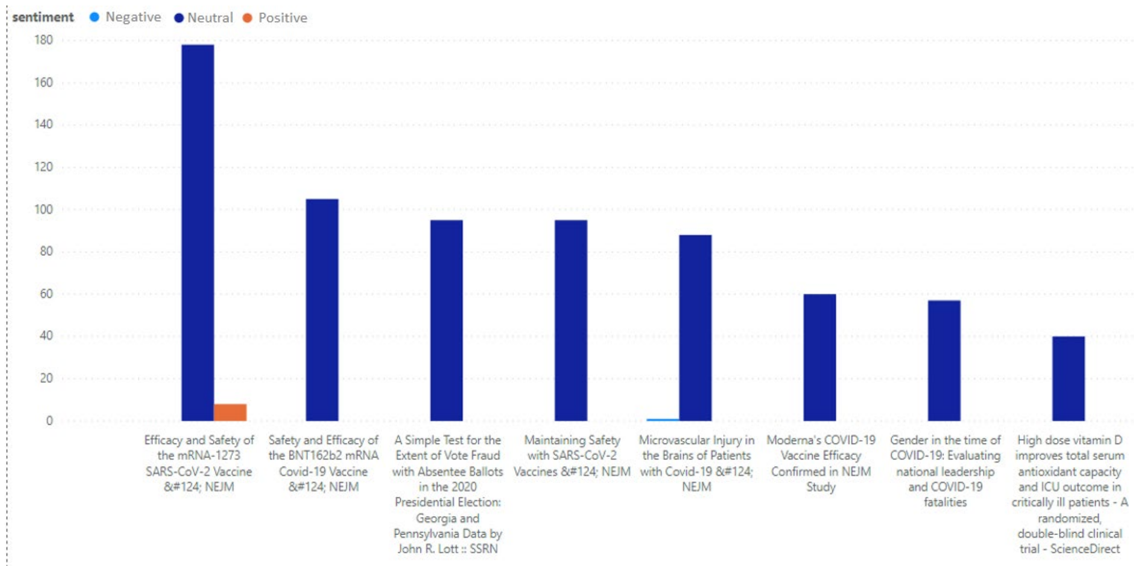

Figure 8. ROs most tweeted and the mention sentiments (with subjectivity variable)

The consideration of the subjectivity variable in the classification model changed the classification of tweets in terms of polarity. Table 3 shows examples of tweets classified as positive. It is verified that no exist negative sentiment classifications.

Table 3. Examples of positives tweets

Positives	▪ RT @andresrguez: The result of Moderna 's vaccine be very very good. http://t.co/VhmUGQ2yaL Via @pmarsupia http://t.co/PaCviqq1x6 ▪ Another awesome application of MPF code: #covid diagnosis through chest CT image processing and #DeepLearning: http://t.co/xm2veVfLMf ▪ RT @robertoguerri: Looks good! Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine | NEJM http://t.co/y8LrjVtBOu

6. Conclusion

This work presents a vision of how social media platforms, in this case, Twitter, can be used to share mentions about ROs. The shares (tweets) were collected and processed, by using Natural Language Processing techniques to extract the sentiments with the following polarity: positive, neutral, and negative.

This development was based on the Python language and libraries. The data validation process required the most effort to reach a level of confidence that guarantees that the data is being treated as intended. One problem identified and solved is the wrong sentiment classification of tweets because there are false positives and negatives, this is solved with the inclusion of the subjectivity variable in a combination of the polarity variable.

The results allow determining the most mentioned ROs, as well, as the sentiment polarity of these mentions. The sentiment polarity of mentions can be seen as recommendations of ROs, which allows for determining the relevance of ROs. Thus, sentiment positive mentions of a RO can be seen as a recommendation, but if the sentiment polarity is negative, it presupposes that the RO has no relevance to be recommended. However, it appears that most mentions are sentiment neutral, this is because they are mentions published by authors or editors announcing the existence of the RO, usually not expressing any opinion/sentiment about it.

It is also verified that there are no negative mentions of ROs, and it can be concluded that it is not normal among the scientific community to share negative comments regarding works published by peers (authors), especially if the authors of these mentions are identified. This is relevant because the peer-review process is normally a blind process, i.e., the review author is not known (shown).

It is more usual to share positive comments on what you like, so comments that were considered positive are really positive, as they convey an opinion and not just a factual presentation of an RO. It appears that there are many shares on Twitter, but most are a mere reference to a RO and do not present any kind of sentiment/opinion (neutral sentiment). This research can serve as a basis for a RO recommendation system based on sharing mentions on social media platforms. A mention, even a neutral mention, is a reference (recommendation) to a RO and if it is positive then it makes that RO even more relevant.

This work has limitations, namely the fact that only one social media platform was used, but as mentioned above, access to data from social platforms is increasingly limited.

In future work, it is suggested to study who makes mentions on social networks to identify and understand who are the science influencers and what types of recommendations they make.

Acknowledgments

This work has been supported by IViSSEM: POCI-01-0145-FEDER-28284.

References

Al-Shamri, M., User profiling approaches for demographic recommender systems, *Knowledge-Based Systems*, 100, pp. 175–187, 2016. https://doi.org/10.1016/j.knosys.2016.03.006

Arora, A., Bansal, S., Kandpal, C., Aswani, R., and Dwivedi, Y., Measuring social media influencer index- insights from Facebook, Twitter and Instagram, *Journal of Retailing and Consumer Services*, vol. 49, pp. 86–101, 2019. https://doi.org/10.1016/j.jretconser.2019.03.012

Babin, J., and Hulland, J., Exploring online consumer curation as user-generated content: A framework and agenda for future research, with implications for brand management, *Spanish Journal of Marketing - ESIC*, vol. 23, no. 3, pp. 325–338, 2019. https://doi.org/10.1108/SJME-07-2019-0053

Bai, X., Wang, M., Lee, I., Yang, Z., Kong, X., and Xia, F., Scientific paper recommendation: A survey, *IEEE Access Institute of Electrical and Electronics Engineers*, 2019. Inc. https://doi.org/10.1109/ACCESS.2018.2890388

Bamakan, S., Nurgaliev, I., and Qu, Q., Opinion leader detection: A methodological review, *Expert Systems with Applications*, Elsevier Ltd., 2019. https://doi.org/10.1016/j.eswa.2018.07.069

Bik, H., and Goldstein, M., An Introduction to Social Media for Scientists, *PLoS Biology*, vol. 11, no. 4, 2013. https://doi.org/10.1371/journal.pbio.1001535

Bornmann, L., Scientific peer review, *Annual Review of Information Science and Technology*, 45, 2011. https://doi.org/10.1002/aris.2011.1440450112

Byrne, E., Kearney, J., and MacEvilly, C., The Role of Influencer Marketing and Social Influencers in Public Health, *Proceedings of the Nutrition Society*, 76, 2017. https://doi.org/10.1017/s0029665117001768

Deng, F., Utility-based Recommender Systems Using Implicit Utility and Genetic Algorithm, *Atlantis Press*, pp. 860–864, 2015. https://doi.org/10.2991/meic-15.2015.197

Devchand, R., Sheehan, P., Gallivan, J., Tuncer, D., and Nicols, C., Assessment of a National Diabetes Education Program diabetes prevention toolkit: The D2d experience, *Journal of the American Association of Nurse Practitioners*, 29(9), 514–520, 2017. https://doi.org/10.1002/2327-6924.12499

Dhanesh, G., and Duthler, G., Relationship management through social media influencers: Effects of followers' awareness of paid endorsement, *Public Relations Review*, vol. 45, no. 3, 101765, 2019. https://doi.org/10.1016/j.pubrev.2019.03.002

Ewing, M., 71% more likely to purchase based on social media referrals, *HubSpot*, United States, 2011.

Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., and Kashef, R., Recommendation systems: Algorithms, challenges, metrics, and business opportunities, *Applied Sciences (Switzerland)*, vol. 10, no. 21, no. 1–20, 2020.

Jomsri, P., Sanguansintukul, S., and Choochaiwattana, W., A framework for tag-based research paper recommender system: An IR approach, *In 24th IEEE International Conference on Advanced Information Networking and Applications Workshops*, WAINA 2010, pp. 103–108, 2010.

Kartajaya, H., Kotler, P., and Hooi, D., Marketing 4.0: Moving from Traditional to Digital, *In Asian Competitors,* pp.

99–123, World Scientific Publishing Co. Pte. Ltd, 2019.

Keith, N., Lettura, E., and Villegas, C., Sentiment analysis and opinion mining applied to scientific paper reviews, *Intelligent Data Analysis*, vol. 23, no. 1, pp. 191–214, 2019.

Ki, C., Cuevas, L., Chong, S., and Lim, H., Influencer marketing: Social media influencers as human brands attaching to followers and yielding positive marketing results by fulfilling needs, *Journal of Retailing and Consumer Services*, 55, 102133, 2020.

Klampanos, I., Manning, C., Raghavan, P., and Schütze, H., Introduction to information retrieval, *Information Retrieval*, vol. 12, no. 5, pp. 609–612, 2009.

Montgomery, D., *Introduction to Statistical Quality Control*, 6th Edition, John Wiley & Sons, New York, 2009.

Morris, M., and Ogan, C., The internet as mass medium, *Journal of ComputerMediated Communication*, vol. 1, no. 4, pp. 1–10, 1996.

Nascimento, C., Laender, A., Silva, A., and Gonçalves, M., A source independent framework for research paper recommendation, *In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, pp. 297–306, 2011.

Nielsen, *Sales Effect Study: Influencer Marketing*, TapInfluence, 2016.

Olenski, S., *Are Brands Wielding More Influence In Social Media Than We Thought?*, Forbes, 2012.

Parra-Santander, D., and Brusilovsky, P., Improving collaborative filtering in social tagging systems for the recommendation of scientific articles, *In Proceedings - 2010 IEEE/WIC/ACM International Conference on Web Intelligence*, WI 2010, Vol. 1, pp. 136–142, 2010.

Rakuten, *Influencer Marketing - Global Survey - Consumers*, Rakuten Marketing, 2019. Retrieved from https://www.iab.com/wp-content/uploads/2019/03/Rakuten-2019-Influencer-Marketing-Report-Rakuten-Marketing.pdf

Ramzan, B., Bajwa, I., Jamil, N., and Mirza, F., An intelligent data analysis for hotel recommendation systems using machine learning, *In Scientific Programming*, 2019.

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J., Recommender Systems for Large-scale E-Commerce : Scalable Neighborhood Formation Using Clustering, *In Proceedings of the fifth international conference on computer and information technology*, 2002.

Schafer, J., Frankowski, D., Herlocker, J., and Sen, S., Collaborative filtering recommender systems, *In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 4321 LNCS, pp. 291–324, 2007. Springer Verlag.

Sen, M., and Udgirikar, S., Survey Paper on Web Recommendation System, *International Journal of Science and Research*, 4, pp. 2319–7064, 2013.

Silva, L., *Cognitive and social implications of the globalization of telematic networks and services: study of the implications of reticular communication in the cognitive and social dynamics of the Portuguese Scientific Community*, Centro Atlântico, 2002.

Sugiyama, K., and Kan, M., Serendipitous recommendation for scholarly papers considering relations among researchers, *In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, pp. 307–310, 2011.

Wang, G., He, X., and Ishuga, C., HAR-SI: A novel hybrid article recommendation approach integrating with social information in scientific social network, *Knowledge-Based Systems*, vol. 148, 85–99, 2018. https://doi.org/10.1016/j.knosys.2018.02.024

Watermeyer, R., Social network science: Pedagogy, dialogue, deliberation, *Journal of Science Communication*, vol. 9, no. 1, pp. 1–9, 2010.

Wielki, J., Analysis of the role of digital influencers and their impact on the functioning of the contemporary on-line promotional system and its sustainable development, *Sustainability (Switzerland)*, vol. 12, no. 17, pp. 7138, 2020.

Xu, L., Jiang, C., Chen, Y., Ren, Y., and Liu, K., User participation in collaborative filtering-based recommendation systems: A game theoretic approach, *IEEE Transactions on Cybernetics*, vol. 49, no. 4, pp. 1339–1352, 2019.

Biography

Professor Jorge Oliveira e Sá joined the University of Minho in September 1999. He holds a Ph.D. in Information Technologies and Systems, with a master's degree in Information Systems Management, from the University of Minho. Currently, he is an Assistant Professor in the Information Systems Department of the School of Engineering of this university, where he has been teaching topics related to Business Intelligence (BI), Data Engineering, Information Visualization, Data Bases, and Information Systems Integration, in several undergraduate and postgraduate courses in the area's ICT and engineering. As a researcher, he is a member of the ALGORITMI Research Center, working in the areas of Business Intelligence, Business Process Management (BPM), Data Analytics, Information Visualization, and Big Data, with specific interests in organizational decision support solutions, business process improvement, and Big Data solutions. Integrates the Program Commissions of several scientific, national, and international conferences. He has advised several master's students.