

1. COVID-19, TEXT AND DATA MINING AND
COPYRIGHT: THE BRAZILIAN CASE *

Allan Rocha de Souza**

ABSTRACT

The COVID-19 pandemic has intensified the importance

requiring the employment of special analysis techniques to obtain results 'by approximation' (e.g., filtering, selection, and sampling). The main difference is that today, with the improvement of information and communication technologies, it has become possible to analyze vast amounts of data in real-time and obtain

creating simple rules for acknowledging the original
creators-10 original

into a single location and converted into an intelligible format. Then TDM is applied to find correlations and patterns from which a variety of information can be extracted and evaluated by the analyst or by the machine, and then presented to the user.¹⁵

With the increase in computing power and the ever-growing production of data seen today, it follows that the amount of data one has to work with usually demands processing power that far exceeds what human beings can achieve, making it unfeasible to operate a large database or mine its contents without proper assistance. The use of artificial intelligence (AI) systems is to assist in these tasks – instead of letting the AI systems perform entirely on their own – is therefore on the increase, through a so-called machine learning process. Machine learning refers to the use of procedures and techniques that enable a machine to process information and, as the name suggests, learn from it, extracting information that will serve as a basis for task-solving and pursuing assigned objectives in a flexible way.¹⁶

The use of machine learning has been advocated to address several issues, including the COVID-19 pandemic. Recent publications propose using AI systems in the formulation of predictive models of case growth¹⁷ as well as for developing new medication¹⁸ and diagnosing methods for COVID-19, be it through information analysis¹⁹ or imaging.²⁰ In Brazil, Aos Fatos demonstrates that, when vast quantities of data are produced and replicated every second, sorting through what is true or

false as well as signaling incorrect content that reappears regularly, can be quite a challenge for a human team of experts to do consistently. The use of AI algorithms, e.g., the ones operating the Radar has become increasingly important in fact-checking.

On the other hand, while AI and algorithms are generally capable of rapidly processing unimaginable amounts of data, they do not have the same sophistication of thought that humans possess. Machine-learning tools usually require millions of inputs in order to apprehend simple information that an ordinary person would be able to gather in no more than a glance.²¹ The processing and classification of large amounts of data are, therefore, crucial for an algorithm to learn and achieve the intended results of its operation.²² Drawing on the work of Russel and Norvig,²³ it is possible to assert that data is a central element today in the operation and training of AI systems, assuming its position as a key player in data analysis which wh0319 Tc[* nQq0.000008871 0 595.32 841.92 reW* nBT/F1

¹⁵ Han J, Pei J, Kamber M, (Elsevier 2011) 7; Kroenke DM, (8th ed., Pearson 2016) 493; Kelleher J, Tierney B, (MIT Press 2018) 241-242.

¹⁶ Samuel AL, 'Some Studies in Machine Learning Using the Game of Checkers' (2000) 4 (1/2) IBM Journal of Research and Development, 207-226; Kaplan A, Haenlein M, 'Siri, Siri, in my hand: Who's the fairest of the land? On the interpretations, illustrations, and implications of artificial intelligence' (2018) 62 Business Horizons, 15-25.

¹⁷ Marni M, Chokani N, Abhari RS, 'COVID-19 Epidemic in Switzerland: Growth Prediction and Containment Strategy Using Artificial Intelligence and Big Data' (2020) medRxiv <<https://www.medrxiv.org/content/10.1101/2020.03.30.20047472v2>> accessed 20 June 2021.

¹⁸ Ho D, 'Addressing COVID-19 Drug Development with Artificial Intelligence' (2020) 2 (5) Advanced Intelligent Systems <<https://onlinelibrary.wiley.com/doi/full/10.1002/aisy.202000070>> accessed 20 June 2021.

In (d) (2) of the FOIA, databases can be defined as structures created out of information collected from various sources. In practical terms, their purpose is to allow the collected data to be preserved and accessed in a more organized fashion, enabling various pieces of information to be effectively cross-referenced. Such organization includes the creation and storage of metadata and indexes, as well as descriptions of the applications used. These groupings or collections of data and metadata are part of a large system that includes computer software dedicated to creating, processing, and administering these databases – which comprise the database management system (DBMS): the applications that work as an interface between users and the DBMS, and the users themselves.²⁵

Databases play a key role in the architecture of data collection and analysis, serving two primary purposes: first, as points of origin for extraction, query, and operational activities, and secondly, as data warehouses i.e. points of destination where all the data collected from multiple sources are reunited so that its contents can be adequately modeled and prepared for the upcoming analysis, and the results of data mining can be stored for future cross-referencing.²⁶ The assembly, acquisition, and maintenance of large databases are, therefore, vital components within the framework that is built to make data mining possible.

4.

It is also important to note that, while some projects can rely on relatively open databases widely adopted by fellow researchers (such as in the COVID-19 genetic sequencing case) or do possess human and financial resources to carry out fast and intensive data-gathering work on their own (such as 'Fato ou Fake' and 'Aos Fatos'), not all research projects fall under these two possibilities. In fact, in the a





of healthy political discourse and the improvement of social welfare throughout the world.

As a final suggestion, we take the opportunity to advise developing countries to incorporate, as soon as possible, friendlier and more expansive limitations and exceptions to copyright that could ensure the proper balance under current circumstances as well as promote one's ability to innovate, such as permitted uses for research purposes and TDM activities. Furthermore, as has become clear, this is not simply a local and national problem, but an international one, as such activities are increasingly carried under cross-border cooperation. This issue may call for international agencies such as World Intellectual Property Organization to take a fresh look into the need to promote worldwide research and text and data mining limitations and exceptions, including by means of discussion and adoption of legal instruments, model legislation and guidance documents

BIBLIOGRAPHY

Abhari RS, Marini M, Chokani N, 'COVID-19 Epidemic in Switzerland: Growth Prediction and Containment Strategy Using Artificial Intelligence and Big Data' (medRxiv 2020).

Alvarenga MB, 'MineraçãC ¢ B

(2020)

<<https://web.archive.org/web/20210309015721/http://www.blog.saude.gov.br/index.php/perguntas-e-respostas/54104-confira-a-entrevista-sobre-o-sequenciamento-do-coronavirus>> accessed 9 June 2021.

Brown K, 'Digital Rights Management: Trafficking in Technology That Can Be Used to Circumvent the Intellectual Property Clause' (Houston Law Review 803 2013) 804.

Carroll MW, 'Copyright and the Progress of ScienceD



