http://dx.doi.org/10.18623/rvd.v15i31.1142

BIG DATA, MACHINE LEARNING AND ENVIRONMENTAL PRESERVATION: TECHNOLOGICAL INSTRUMENTS IN DEFENSE OF THE ENVIRONMENT

Carlos Alberto Molinaro

Professor at the Faculty of Law of the Pontifical Catholic University
of Rio Grande do Sul - PUCRS
Email: carlos. molinaro@pucrs. br

Augusto Antônio Fontanive Leal

Doctorate in Law from Pontifical Catholic University of Rio Grande do Sul (PUCRS). Master
and graduated in Law from the University of Caxias do Sul (UCS).
Email: augustoafleal@gmail.com

ABSTRACT

The purpose of this article aims a descriptive analysis of the contents of new technologies called “Big Data” and “machine learning” and, with these, make an approximation of the technologies, as environmental protection tools. The research will be divided into two essential topics: a) the approach and conceptualization of technological resources of Big Data, and the techniques of machine learning; b) an exposition of the improvement resulting from the use of Big Data, and of machine learning techniques, with the purpose of environmental preservation. By means of this, it will be possible to prove, from feasible examples, how the use of technologies can influence and demonstrate a new direction for environmental preservation. Reinforcing the duty of the State and Society in the protection and defense of nature. The research was qualitatively supported by the hermeneutic method with the use of bibliographical and documentary research.

Keywords: Big Data; *machine learning;* environmental preservation; environment.

BIG DATA, MACHINE LEARNING E A PRESERVAÇÃO AMBIENTAL: INSTRUMENTOS TECNOLÓGICOS EM DEFESA DO MEIO AMBIENTE

RESUMO

O presente artigo tem por objetivo realizar uma análise explicativa sobre o conteúdo das novas tecnologias denominadas de Big Data e machine learning e, com isso, efetuar uma aproximação da utilização destas tecnologias como ferramentas de proteção ambiental. Para isso, a pesquisa se dividirá em dois tópicos essenciais: a) a abordagem e conceituação dos recursos tecnológicos do Big Data e das técnicas de machine learning; b) uma exposição do aprimoramento advindo da utilização do Big Data e de técnicas de machine learning para a finalidade de preservação ambiental. Por meio disso é que se poderá comprovar, a partir de exemplos factíveis, como a utilização de tecnologias como Big Data e machine learning podem influenciar e demonstrar um novo rumo para a preservação ambiental, reforçando o dever do Estado e da Sociedade na proteção e defesa da natureza. A pesquisa será qualitativa amparada pelo método hermenêutico com uso de pesquisa bibliográfica e documental.

Palavras-chave: *Big Data; machine learning; preservação ambiental; meio ambiente.*

INTRODUCTION

Due to the constant impacts generated on biodiversity in a global scenario, it is already possible to foresee an ecological calamity *in fieri*. The complexity of environmental problems is entangled in issues that are difficult to understand, particularly with regard to environmental risks. Likewise, issues related to the principles of prevention and precaution are also difficult access to the different dependencies found in biodiversity, based on the structure of the environment and the various relationships and consequences triggered by each human act. Nonetheless, it seems that with the use of technological resources it may be possible to achieve more efficient means of combating environmental degradation not only at local levels, but globally. In this sense, the uncontested exponential evolution of technology has allowed nowadays to be able to count on the good practices of Big Data and *machine learning.*

It is through this study that we will seek to establish a relationship between large technological innovations, such as Big Data and *machine learning* algorithms*,* making environmental protection tools more accurate that only a large amount of data and a portentous processing power could offer. To this end, the present essay is divided into two chapters. In the first one, an approach will be tried on the method of Big Data and the techniques of *machine learning*, propitiating its understanding. In the second, we will try to approximate these technological innovations in the proposal of environmental preservation, with the study of existing programs that use these technologies to increase and improve nature conservation.

1 BIG DATA AND *MACHINE LEARNING*: APPROACH AND CONCEPTS

In a connected world, data is constantly generated. There are a number of technological devices that enable this growing generation of data and make it possible to establish standards, a comprehensive understanding of the world, or even the predictability of events. With the exponential technological growth, the speed of processors and the large-scale storage capacity, the massive figure of Big Data and the processing of this data was made possible, allowing an understanding that, without *machine learning* techniques, would not be possible. To deal with these issues, this chapter will be divided into two sections: the first, in which will be studied the Big Data and the opportunities that are provided by its unceasing growth; the second, in which will be studied the cognitive revolution by means of *machine learning* techniques in order to be able to find patterns in the enormous amount of data generated, allowing the understanding of this technological tool in development.

1. 1 Big Data: looking at a new world of opportunity

Of course, the use of data has been enhanced by the sophistication offered by technology and its incessant advances, but it is not a task due to the emergence of computers. Patterns have long been sought in data collected, such as the search carried out by a farmer by patterns in a culture, or a politician in the opinion of a voter (WITTEN; FRANK; HALL, 2011, p. 4). Data have always been stored in some way in the human experience, for the most diverse questions, whether in an individual aspect or in memory, or in documents and records. If at a given historical moment the data storage was performed by a hunter seeking to understand a pattern in the migratory behavior of his hunt, this may have gone into the storage of data to facilitate trade.

However, in a technological development environment, one can say that data overload is occurring. As Witten, Frank and Hall refer, the amount of data in the world and in human life is growing and there is no end to it (WITTEN; FRANK; HALL, 2011, p. 3). According to the authors, this is due to the omnipresence of computers in everyday life, which, through not-so-expensive storage sites, such as online storage, bring with it the ease of saving what was previously dispensed (WITTEN; FRANK; HALL, 2011, p. 4). The data is simply being stored over the days, by virtual means, in a scenario which at any time the memory of any device can be amplified to hold a greater number of data. However, the generation of data is not limited to the common life of every human being, but it is in countless corporate, governmental, economic and social activities.

Demonstrating the processor development capability (CPU) still in 1965, Gordon Moore proved that the number of *transistors per square* would double each year, and in 1975, revising his theory, Moore doubled that time frame to two years (BELL, 2016). Taken together, there is a decline in the cost of storing data (KATZ, 2013, p. 916), providing a fertile ground for the development of what is coined Big Data.

How big is Big Data?

Katz refers to a subjective and ever-moving target, so that one of the best ways to understand this concept would be to look back (KATZ, 2013, p. 916). Continuous enhancement of memory systems coupled with increasingly expanded processing capacity demonstrates that on a *continuum,* there is a storage possibility that is simply unable to be understood in ordinary, isolated electronic devices.

However, *big data* remains an abstract concept, so that in spite of the wide recognition of its importance, differing opinions about its definition depend, and can be conceptualized, in general, as *datasets* that cannot be perceived, acquired, administered and processed by traditional software and hardware within a tolerable time (CHEN et al., 2014, p. 2). According to the International Data Corporation (IDC) definition, Big Data describes a new generation of technologies and architectures, designed to economically extract values from huge volumes of a wide variety of data, through a highly rapid capture, discovery and/or analysis (GANTZ; REINSEL, 2011, p. 6). Therefore, it is important to clarify that Big Data is not only about the original content stored or consumed, but also about the information involved in this consumption, which can be exemplified by the way *smartphones* produce a source of additional data being captured, such as is the case of the geographical location, text messages and history of Internet browsers (GANTZ; REINSEL, 2011, p. 6).

Generated by the most diverse technological means, the data in an absurdly difficult quantity to be perceived, apprehended and measured, do not attending a linear growth, but rather exponential. As explained by Katz, there is an exponential development in a synergy between the concepts of Moore's Law, Big Data, and the artificial intelligence revolution, so that each doubling of the speed of processors, of dividing by half the price for data storage and the advances of *machine learning* open up an immensity of possibilities in a non-linear perspective (KATZ, 2013, p. 922).

In this horizontally transverse scenario, Big Data is formed from several sources and may include varied types of data in an amount difficult to be absorbed. There is a great deal of new data sources present in social media where users, sometimes even unconsciously, end up generating a large amount of data that can be used in a variety of ways and that are expanding at fast algorithmic rates. As Gantz and Reinsel (2011, p. 6) describe:

> Big data is a horizontal cross-section of the digital universe and can include transactional data, warehoused data, metadata, and other data residing in ridiculously large files. Media/entertainment, healthcare, and video surveillance are obvious examples of new segments of big data growth. Social media solutions such as Facebook, Foursquare, and Twitter are the newest new data sources. Essentially, they have built systems where consumers (consciously or unconsciously) are providing near continuous streams of data about themselves, and thanks to the "network effect" of successful sites, the total data generated can expand at rapid logarithmic rates.

However, while social media is a major contributor to data generation, it is not the only source. In addition to the social media platforms that produce a large amount of data through the networks of individual profiles or other content produced, which deal with blogs, user discussion forums, even "*likes",* contains potential information to be used, and may be referred to as sources the data produced by activities as the mechanically generated content in the form of device log files, in the internet function *of things*. Other sources of data may still be found in software services, cloud applications and the hitherto unexplored ones, are medical record data and customer correspondence (RUNCIMAN, 2014, p. 1). Not surprisingly, according to IBM (2017), 2. 5 quintillions of bytes of data are created every day. It is so much that 90% of the data in the world today has been created in the last two years, dealing with data that came from everywhere: climate information sensors, social media posts, digital photos, videos, transactional records, GPS signals from cell phones, among countless other ways.

The possibilities available from the use of Big Data are the most diverse and represent a fertile field for the improvement of human life and all species of life on earth. It is necessary to have the means to use this data, which includes its identification, verification, understanding and use. Basu and Hall, demonstrating the ability to use this technological feature, cite the presence of Big Data in the *Square Kilometer Array*, a radio telescope with a size that can make it fifty times more sensitive and ten thousand times faster than any other, which will generate by day the amount of data on the internet, making astronomers able to perceive any signal in the galaxy and look deeper into space than ever before (BASU, HALL, 2014, p. 46); according to the same authors, Big Data is part of a revolution in the domestic, social, and business world, for just as the massive *Square Kilometer Array* radio telescope will generate Big Data to enable a better understanding of history of man and his place in space, there are great opportunities to increase the use of Big Data to better understand the future human environment and the possibilities on the planet (p. 51).

As a result of an exponential technological development, with increasingly powerful processors and the reduction of the cost of data storage, it was possible to generate a huge amount of data that grows daily, having as a source the creation of data by the conscious or unconscious users of social media or other connected devices, the mechanically recorded of various devices with connection (internet of things), cloud storage and various other data that can be explored. This large-scale data that helps build the meaning of Big Data offers many possibilities and opportunities to positively revolutionize human life on Earth and, as in the case of environmental issues, providing a fertile field for environmental protection.

However, it is not enough to acknowledge the existence of Big Data. It is necessary to enable the data to be used, that is to say, it is necessary to discover or channel the large amount of data by means of which it is possible to use it. As shown, this is an absurdly immense amount of data and there is no way for the human intellect to do this service without a technological tool capable of mining this data, selecting what is being sought and verifying, through some analysis, what is relevant. It is in this sense that *machine learning* and a revolution in the field of artificial intelligence arise, as will be demonstrated in the next subchapter.

1. 2 Cognitive revolution, *machine learning* and data mining

The colossal volume of data that day after day is generated, to be used according to a proposal to achieve an improvement in the human experience by the opportunities that exist and will exist with Big Data depends on a technological apparatus that not only process, and interpret a given amount of data but perform tasks of choice and can effectively learn from the analysis generated to enable behavioral analysis responses by identifying patterns that can be used.

Although tooling of this size can be driven by some human being, it is impossible to conceive that the human mind can work with such a large volume of data as shown in what is coined Big Data. As evidenced by Witten, Frank and Hall, there is a large gap between data generation and the human understanding of these data (WITTEN; FRANK; HALL, 2011, p. 4). It seems certain that as the volume of data continues to increase, the human aptitude to understand this alters, alarmingly, causing a large amount of data to remain unused (WITTEN; FRANK; HALL, 2011, p. 4). These data, with potential utility, rarely come to be identified or used, remaining in a state of invisibility. Data *mining* is not such a recent feature, since economists or communications engineers, for example, have long worked with the idea of finding patterns in data automatically, identifying, validating, and using for predictions. What is new is an astounding increase in the opportunities for these patterns to be found in data that comes from insatiable growth (WITTEN; FRANK; HALL, 2011, p. 4).

Certainly, according to Katz, data storage capacity and processor speed alone are not enough to generate the aggregate set of perceptions that lead to increased productivity and innovation. According to the author, in addition to these two technological innovations, a revolution in artificial intelligence must be observed as a third pillar in a new era of productivity (KATZ, 2013, p. 918). It is in this context that the use of *machine learning* techniques appears (in this text the original term that can be understood in Portuguese as computational learning or automatic learning) appears as a tool capable of finding and describing patterns in data (WITTEN; FRANK; HALL, 2011, p. 5).

However, there is a metaphorical content in saying that a computer can learn, appearing a merely illustrative expression. Analyzing five definitions about learning: a) having knowledge of something through study, experience, or learning; b) become aware of an observation or information; c) memorize; d) be informed or find out something; e) receive instruction, ask Witten, Frank and Hall about how one could know if a computer has learned something? Or if a computer became aware of something? The question that concerns the possibility of a computer being aware or conscious is a current issue of philosophy. However, the last three aspects (items "c", "d" and "e") could be very close to the meaning of *machine learning* (WITTEN; FRANK; HALL, 2011, p. 7). In that case, instead of seeking a replica of the human cognitive process in computers that had been based on the belief of creating artificial versions of human brain functionality, being bound to a supposed intelligent process of production, is the alternative possibility of obtaining results that automated processes reach, being precise, appropriate, auxiliary and useful, they can be considered intelligent, even if they do not come from an artificial replica of a human cognitive process (SURDEN, 2014, p. 98-99).

In this way, learning could be attributed to things when they change their behavior in a way that can perform better in the future, being much more a question of performance than knowledge (WITTEN; FRANK; HALL, 2011, p. 7). As mentioned by Surden, operating well, *machine learning* algorithms can produce approximate automatic results of what would be done by an equally situated person (SURDEN, 2014, p. 90). So, the use of data, by mining and subsequent verification in order to find a pattern of behavior that can be used not only to understand something that has occurred or is occurring, but to dispose of a forecast in the face of a great deal of data (Big Data), would be given by *machine learning* techniques.

With the example of automatic classification of emails as spam, Surden seeks to demonstrate that *machine learning* algorithms can develop more and more to receive a greater amount of data, aiming at the construction of a model of a complex phenomenon that will enable a computer to perform accurate classifications (SURDEN, 2014, p. 92). Based on this, it can be said that computers can have "smart" results in tasks without the intervention of human cognition (SURDEN, 2014, p. 95).

In short, by the learning process one would have the improvement of a behavior by the study of experiences (RUSSEL; NORVIG, 2010, p. 693). The techniques of *machine learning* would operate learning from experience and developing an improvement of their performance over time, which would contribute greatly to the extent that it was implemented in the most diverse sectors, such as economic, social, environmental, enabling an understanding and providing forecasts.

It must be said, however, in a problematic aspect present in the forecasts. If the objective of using *machine learning* techniques is based on data analysis to obtain approximate results in the future, in scenarios not yet seen (or never seen), one can run into the problem of generalization (SURDEN, 2014, p. 105). Surden shares the problem of generalization in three aspects: a) if future events depend on single or non-unusual facts, the forecast would be difficult; b) overgeneralization should be avoided, since it is undesirable for *machine learning* algorithms to detect idiosyncratic or biased data that would obviously not be able to predict; c) is the case, for example, of legal predictions, when some *machine learning* algorithm encounters difficulties in capturing subtle factors of great relevance (SURDEN, 2014, p. 105-106).

The problem of generalization could overturn a prediction that was sought when some key factor is not present in the previous data, when there is some case where the data are endowed with false perceptions such as, for example, prejudgments that would prevent a corresponding analysis with reality and, also, when there were subtle factors that would hardly be captured by the technique of *machine learning* used. To act about these factors, it is important that the data used are filtered on criteria such that it can measure certain characteristics. Also, certain problems existing in the possible occurrence of atypically or of difficult perception factors can be improved during the use of developed *machine learning*.

Therefore, in order to maximize the power of *machine learning* techniques, in a computational learning theory based on the intersection between artificial intelligence, statistics and computational science theory, an underlying principle is demonstrated that any hypothesis that is seriously wrong would almost certainly be found after a small number of examples, as would be making a wrong prediction, so any hypothesis consistent with a broad examples system demonstrate an unlikely error and is probably approximately correct (RUSSEL; NORVIG, 2010, p. 714). To do so, the ability to process a large amount of data, testing hypotheses with an increasing approximation of a correct and coherent response, would have the function of eliminating problems arising from a generalization, even when for a forecast it depended on a fact that has not yet occurred, because a series of experiences would lead, after a given period, in the de characterization of this problem by the content of the circumstances hitherto analyzed. In conclusion, Flach conceives *machine learning* as the systematic study of algorithms and systems in order to improve their knowledge or performance through experience (FLACH, 2012, p. 3). Over time, *machine learning* techniques become more profitable and allow, in the face of a great deal of data, an enhancement of various sectors of society, revolutionizing human life.

The use of *machine learning* that, through images generated daily by satellite radars that monitor coastal waters to detect oil stains, allows to train a contamination detection system, with examples of spills for each user for their own purposes, is just an example of how this technology can aid in environmental preservation. In case, when finding possible oil stains, first the system processes and normalizes the image, identifying the suspect spot. Afterwards, the attributes of each oil spot are determined, such as region, size and intensity, and, in the end, pattern learning is used for the resulting attributes (WITTEN; FRANK; HALL, 2011, p. 23).

As stated by Flach, there is a good chance that *machine learning* algorithms are already being used by a large number of people, even though they are not aware of it (Flach, 2012, p. 1). Similarly, Russel and Norvig report the existence of behind-the-scenes performance of artificial intelligence, as in the case of automatic credit card approval. These changes that are incorporated end up modifying the daily life of the people, although many do not realize this. Although it is in development, it can be expected that a medium success of artificial intelligence would affect all the people in their lives (RUSSEL; NORVIG, 2010, p. 1051).

The truth is that the world as it is known has been changing daily with the exponential technological advancement of processors, techniques of memory storage and increase of *machine learning* algorithms. Thus, there is a vast field of possibilities and opportunities to improve human life in several aspects, among which, and certainly one of the most important, is the preservation of nature as a whole. In addition, some techniques have already been developed and serve as examples on how Big Data technology and *machine learning* can bring effectiveness to ensure an ecologically balanced environment, both locally and globally. This will be demonstrated in the next chapter.

2 THE FUNCTION OF BIG DATA TECHNOLOGY AND *MACHINE LEARNING* TECHNIQUES IN ENVIRONMENTAL PRESERVATION

The use of *machine learning* algorithms coupled with a powerful capacity for data storage together represent significant changes in the most varied aspects of human life. Therefore, both Katz (2013) and Surden (2014) seek to relate the use of *machine learning* to the best performance of legal tasks, the first author works in a new era of quantitative legal prediction that would help lawyers, students and faculties of Law (p. 912-913); and the second, warns that it is not insignificant the number of legal tasks that can be partially automated using artificial intelligence techniques (p. 88).

In the quest for the preservation of nature this is no different (!).

The technological development expands the tooling that serves both to find environmental problems and to reach solutions. This perspective may bring greater effectiveness to the fundamental right to the environment and its determinations provided for in Article 225 of the Brazilian Federal Constitution (BRASIL, 1988), as is the case of the defense and preservation of an environmentally balanced environment, with a view to the reality of a res maximi momenti subject that led to the constitutionalization of the foundations of environmental protection and the increase of its quality (MILARÉ, 2014, p. 160-161). However, this concern with the natural reality is not bound up with mere positivation. In addition, the environment reaches the levels of a fundamental right that leads to the formulation of a principle of environmental primarity, recognizing the illicitness of treating the environment as a subsidiary, accessory, minor or negligible (BENJAMIN, 2010, p. 118).

It should be noted that the proposal for a project such as Global Forest Watch (http://www.globalforestwatch.org/), which consists of an online forest monitoring and alert system that provides tools for anyone in the world to be capable of managing and conserving florets, is a clear example of the use of technological tools for environmental preservation. The technological trump of a project of this magnitude is not only based on a gigantic collection of data at global level (Big Data), but also in a union of people of the whole world in the provision of data and use of systems of alert that make possible a greater forest protection.

The aforementioned article 225 of the Federal Constitution imposes, both for the State and for the collectivity, the preservation and defense of the ecologically balanced environment. That is why, according to Sarlet and Fensterseifer, there is a difference between this right and the fundamental rights of first dimension, based on liberal rights, as well as the principle of equality, based on second-degree social rights, concluding by the normative-axiological pillar of the fundamental right to the environment as a principle of duty of solidarity (SARLET; FENSTERSEIFER, 2013, p. 47). With this, the cooperative system itself in high-tech projects using Big Data, *such* as Global Florest Watch, provide means for citizens and the Brazilian government itself to be able to fulfill their duty of environmental protection present in the normative content of the article 225 of the Federal Constitution. In addition to internal regulations, compliance with international treaties dealing with a wide range of global issues must be taken into account, demonstrating that not only effective implementation of international standards, but also the use of Big Data and *machine learning* serve to establish fair treaties and have a precise and necessary regulatory content.

More deeply, however, it will be sought to demonstrate in the following subchapter how significant technological impacts such as Big Data and machine learning techniques can influence and redirect efforts in environmental preservation through a general improvement of the tools to be used.

2. 1 BirdCast and the migration of birds in real time

BirdCast (http://birdcast.info/) is a program designed for the first time to enable predictions of bird migrations, establishing conditions to know when they will migrate, where they will migrate and how much they will fly. The project, developed by The Cornell Lab of Ornithology, proposes that knowledge of migratory behavior will aid in conservation criteria, such as the positioning of wind turbines or even in the identification that at certain night some high building may have to leave its lights on preventing the death of millions of birds. However, the importance of BirdCast goes beyond that. Higuchi, while mentioning having collaborated with American and Asian scientists on the satellite tracking of bird migration for twenty years, demonstrates the importance of migratory bird knowledge for the conservation of the global environment (HIGUCHI, 2012, p. 3). The author points out that migratory birds encounter various problems during migration, including habitat destruction, chemical pollution, hunting, and collisions with airplanes (HIGUCHI, 2012, p. 3).

There is a link between migratory birds and future issues about the relationship between the environment and human activities. Every fall and spring a large number of birds visit several areas on their migratory routes and, as without food there is no migration, you get food and other significant issues in these areas. Consequently, migratory birds play an important role in maintaining ecosystems in different areas due to their eating habits, so if only a few migratory birds exist, the insect population could increase dramatically in forests causing an imbalance in nature, as well as algae could grow densely in lakes, polluting water (HIGUCHI, 2012, p. 11). Sometimes these negative impacts occur when a change in bird migration is not well understood. Declining populations of migratory birds in a given habitat may lead to deterioration of an ecosystem in another area. According to Higuchi, the destruction of tropical forests in Southeast Asia decreases the number of tropical migratory birds that spend winter there and migrate to Japan in the spring, creating an imbalance in the nature of forest ecosystems in Japan. Also, the destruction of tidal flats in Japan, may decrease the number of many shorebirds, causing a deterioration of wetland ecosystems in the Philippines, Australia and Russia (HIGUCHI, 2012, p. 12). Therefore, issues such as these represent a challenge for the development of ornithology and ecology, evidencing that the conservation of birds is not only for their preservation, but also for the maintenance of ecosystems, leading to the conservation of the global environment (HIGUCHI, 2012, p. 12).

The BirdCast program provides accurate models of bird migration, allowing researchers to understand behavioral aspects of migration, such as migration time and migration pathways responding to climate change, and which links exist between migration time and subsequent size changes in populational dimensions. Among the data sources of the program are the flight calls that allow to identify the composition of nocturnal migration species when it occurs providing almost real time information, meteorological surveillance radars, that can be used to observe and track the movements of birds and, finally, the revolutionary eBird.

Started in 2002 by The Cornell Lab of Ornithology and the National Audubon Society, eBird is a program that provides a real-time list to report reports and access information about birds, providing data sources for basic information on bird abundance and distribution in spatial and temporal scales. The eBird program is one of the largest and fastest accumulators of biodiversity data resources out there, having in May 2015 more than nine million reports of bird observations across the world. Drawing on these data sources, including the portentous feature of eBird, the BirdCast program proposes the development of two *machine learning* techniques, Collective Graphical Models (CGMs) and Semi-Parametric Latent Process Models (SLPMs). The two techniques are designed to reconstruct and predict the behavior of four hundred species of migratory birds in North America, resulting in a model capable of identifying the complex conditions that govern the dynamics of migratory behavior. In addition, BirdCast seeks the development of interoperable data to synthesize bird observations, flight calls, radar data and data from multiple sources such as satellite imagery, meteorology, and human population data. It counts, in the end, with a web data system to communicate migratory forecasts generated by BirdCast to the public.

With the use of *machine learning* techniques in conjunction with a huge cumulative data source, BirdCast enables the understanding of migratory bird behavior, providing valuable information to significantly increase environmental preservation on a global scale.

2. 2 Global logistics chain and the identification of key points of endangered species

The logistics chain represents the trajectory of a product, from conception as a supply, passing through the stage of production in a process of industrialization and reaching, in the end, the distribution, that meets the demand of consumers. Not infrequently, this process takes place on a globalized level, generating an interconnection of several parts in this complex process of production. Companies that act on critical areas for endangered species, however, tend to continue with over-exploitation further aggravating the problem portrayed.

Loss of biodiversity and, more precisely, local threats to species, are being driven by economic activity and consumer demand (LENZEN et al., 2012, p. 111). For this reason, 30% of species around the world are threatened by international trade (LENZEN et al., 2012, p. 109). Understanding this shows that there is no way to act only at the local level, around a regulation that does not only affect developing countries that export in large numbers, there is a need for regulations to reach countries where consumption actively acts in the creation of a heavy environmental footprint placing numerous species in threat.

The issue is precisely in the scenario that involves the identification of central points of threatened species from a global logistic chain. And it is precisely this trump of a research conducted based on the construction of a map that combines several maps of occurrence of a series of endangered species. With these spatial data from an environmental perspective, it is possible to know which countries and which categories of consume threaten habitats at various critical points (MORAN; KANEMOTO, 2017, p. 1). The research has brought remarkable results, such as the environmental footprint of terrestrial species from the United States not only on Southeast Asia or Madagascar, but also at critical points in southern Europe, the Sahel, the east coast and west of southern Mexico, across Central America and Central Asia, and in southern Canada. Despite the increased attention on the Amazon rainforest, the study revealed that the United States footprint in Brazil is higher in the plateau to the south than in the Amazon. The deep US footprint on biodiversity in southern Spain and Portugal in relation to the impact on the number of endangered species of birds and fish is noteworthy as these areas are rarely considered as a critical point of species threatened (MORAN; KANEMOTO, 2017, p. 1). Through the research it was also possible to observe critical points of threats for several major consumer countries and to approach in each particular region affected by consumption. This is the case of the impact of US consumption in South Africa, Europe's consumption in Africa, and Southeast Asia by Japanese consumption (MORAN; KANEMOTO, 2017, p. 1).

Through the use of the critical points method, it has become possible to identify areas where the threat to biodiversity is predominantly driven by a small number of countries. By identifying the regions in which certain countries are pushing for biodiversity, it is possible to initiate direct collaboration between producers and consumers, taking into account international regulations, in order to mitigate the environmental impacts in the identified sites (MORAN; KANEMOTO, 2017, p. 3). By spatially accounting for the explicit biodiversity impacts of certain areas, it is possible to help improve sustainable production, international trade and consumption. Responsibility for environmental impacts must be shared across the entire logistics chain, from industries to final consumers. Since the maps provided by the survey help producers as well as consumers, including conservationists, in the search for solutions to the tracking of biodiversity critical points (MORAN; KANEMOTO, 2017, p. 3).

The research was only made possible by a massive content of data from several sources. The results were calculated from the following three available data (MORAN; KANEMOTO, 2017, p. 5):

1. IUCN Red List of Threatened Species: Contains genetic diversity and ecosystem building blocks, information on conservation and distribution status, and enable informed decisions on biodiversity conservation at local and global level. Provides taxonomic information, conservation status, and distribution for globally evaluated plants, fungi and animals.

2. BirdLife International: This is a global, multi-partner partnership, each of which is a non-profit, non-governmental environmental organization. BirdLife International seeks through rigorous science to implement conservation programs for birds and the whole of nature. The global partnership allows a work for nature in each location where there is an established member, being connected nationally and internationally.

3. Eora MRIO: has a database of *input-outputs* with the aid of social and environmental satellites in 187 countries in 15. 909 sectors.

The importance of this research, based on giant databases, demonstrates the need for a direct approximation of international and local legislation to provide a more accurate preservation of the environment in its wholeness. Identifying the logistical chain and threats to biodiversity occurring in a particular location can lead to the design of negotiations between the various countries involved, seeking shared responsibility and greater conservation of biodiversity.

2. 3 Technological government programs: an example from technological tools used by the US Environmental Protection Agency

According to Breggin and Amsalem, Big Data has been used by government agencies, nongovernmental organizations and private companies to protect the environment, and the authors exemplify it by improving energy efficiency, promoting environmental justice, tracking of climate change and monitoring of water quality (BREGGIN; AMSALEM, 2014, p. 3). It is no coincidence that, according to the authors, in the US federal government, most governmental entities responsible for environmental protection have used Big Data in their work. The use can be illustrated by risk analysis, research, implementation, public education and capacity building.

In the US Environmental Protection Agency, we can mention the air quality monitoring that is being developed in the research and development office, with one of the examples being the focus on new technologies that will help the government to assess the ozone layer and compliance with national air quality standards in conjunction with a data fusion method that combines the results of monitoring to create forecasts (BREGGIN; AMSALEM, 2014, p. 6).

Another example is the development of CompTox, a computational toxicology research program, which aims to address the lack of health and environmental data in thousands of chemicals (BREGGIN; AMSALEM, 2014, p. 7). The program uses advanced research integrating biology, biotechnology, chemicals and computer science to identify important biological processes that can be triggered from chemicals, so that thousands of chemicals can have their potential hazards evaluated in a short space of time.

Another example is The Emissions & Generation Resource Integrated Database (eGRID), an inventory of federal data on power plants and energy companies collected, covering the environmental characteristics of almost all electricity generated in the United States. Among the characteristics are the atmospheric emission of pollutants, the emission rate and the amount of electric energy generated in an electric power plant that is transmitted and distributed to the consumer. These enormous amounts of data are aggregated by facilities, enterprises and the state (BREGGIN; AMSALEM, 2014, p. 7).

With the aim of unifying systems, combining and connecting satellites and ground-based observation systems around the world, The Global Earth Observation System of Systems (GEOSS) is a project under development by The Group on Earth Observations (GEO), an international partnership between governments and organizations. It is intended to assist the international community in coordinating emergency responses to natural or man-made disasters, such as forest fires, monitoring climate change, and managing natural resources (BREGGIN; AMSALEM, 2014, p. 8). Through GEOOS, there is an increased understanding of Earth processes and an improvement in the predictive capabilities that underpin decision making by providing access to data, information and knowledge for a wide variety of users.

The technological approach initiated by the Environmental Protection Agency demonstrates the indispensability of the use of technological resources of high tip for an effective conservation of the environment. The complexity of the various factors that affect biodiversity can be understood by computational resources that have the capacity to store a large amount of data as well as to process the data collected, providing understanding for decisions to be taken at the governmental level. Furthermore, environmental regulation can be more accurate by analyzing a large amount of data, Big Data, in conjunction with the processing of these data, which can predict a number of issues, making it possible to anticipate the occurrence of environmental disasters, among other factors that affect biodiversity.

CONCLUSION

The revolution provided by the ability of computer processors coupled with the considerable increase in data storage resources has led to the current era of Big Data. The speed in data creation in a tremendous amount that makes it impossible to analyze by traditional management systems has increased the possibilities for a greater understanding of several factors in the world, such as economy, society and nature. In conjunction with Big Data, the advent of *machine learning* techniques, enabling a processing of this vast amount of data, enables this understanding, providing a technological tool like never before and bringing in the challenges of developing and improving this technology, innumerable possibilities for sectors of society.

In this essay, we sought to approach the Big Data era as a way to protect and preserve the environment. The interconnection of biodiversity in several regions demonstrates the diffuse character of the environmental good and the understanding of its webs of connections through a large amount of data that surpasses the local barriers and allows an approximation of the facts as they occur at the global, state, regional and local levels. With this, a fair dialogue is possible between the parties involved to establish an equivalent responsibility for environmental preservation, as well as bringing matters to be regulated at the international and internal levels of each country.

Existing projects, be they nongovernmental members from around the world in cooperation for environmental preservation, whether governmental, as exemplified by developments in the US Environmental Protection Agency, demonstrate how Big Data has been assisting government in preservation of nature, reaching the post of key piece for more precise decisions. In conclusion, the use of Big Data and *machine learning* techniques, besides being possible, is already a system used to preserve the environment; it can be affirmed that in the Brazilian internal plane, the use of these technological resources must be implemented with the purpose of bringing effectiveness to the fundamental right to the environment ecologically balanced, in the political decision-making, in the public policies of environmental justice, in the jurisdictional exercise and in the most diverse fields of state performance.

REFERENCES

BASU, Dalim; HALL, Jon G. *Big data, big opportunities*. In: BCS, The Chartered Institute for IT. *Big data: opportunities and challenges***.** Swindon: BCS Learning and Development Ltd., 2014.

BELL, Lee. *What is Moore's Law? WIRED explains the theory that defined the tech industry*.

WIRED, ago. 2016. Disponível em: < http://www.wired.co.uk/article/wired-explains-moores-law>. Acesso em: 19 abr. 2017.

BIRDLIFE INTERNATIONAL - *Página Web*. Disponível em: <http://www.birdlife.org/worldwide/partnership/about-birdlife>. Acesso em: 02 mai. 2017.

BENJAMIN, Antônio Herman. *Direito Constitucional Ambiental brasileiro*. In:

CANOTILHO, José Joaquim Gomes; LEITE, José Rubens Morato (orgs.). *Direito constitucional ambiental brasileiro***.** São Paulo: Saraiva, 2010.

BIRDCAST. *Página Web*. Disponível em: <http://birdcast.info/>. Acesso em: 30 abr. 2017.

BRASIL. *Constituição da República Federativa do Brasil de 1988*. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm>. Acesso em: 24 abr. 2017.

BREGGIN, Linda; AMSALEM, Judith. *Big Data and Environmental Protection: An Initial Survey of Public and Private Initiative***s.** Washington: Environmental Law Institute, 2014.

CHEN, Min; MAO, Shiwen; ZHANG, Yin; LEUNG, Victor C. M. *Big* *data: related technologies, challenges and future prospects.* New York: Springer, 2014.

EBIRD. *Página Web*. Disponível em: < http://ebird.org/content/ebird/>. Acesso em: 30 abr. 2017.

EORA MRIO. *Página Web*. Disponível em: <http://www.worldmrio.com/>. Acesso em: 02 mai. 2017.

FLACH, Peter. *Machine Learning: the art and Science of algorithms that make sense of data***.** Cambridge: Cambrige University Press, 2012.

GEO. *Página Web*. Disponível em: <http://www.earthobservations.org/geoss.php>. Acesso em: 02 mai. 2017.

GLOBAL FOREST WATCH. *Página Web*. Disponível em: <http://www.globalforestwatch.org/>. Acesso em: 24 abr. 2017.

HIGUCHI, H. *Bird migration and the conservation of the global environment*. *Journal of Ornithology,* v. 153, n. 1, p. 3-14, 2012.

IBM. *What is big data?* Disponível em: <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html> Acesso em: 21 abr. 2017.

KATZ, D. M. *Quantitative legal prediction—or—how i learned to stop worrying and start preparing for the data-driven future of the legal services industry*. Emory Law Journal, v. 62, n. 1, p. 909-966, 2013.

LENZEN, M.; MORAN, D.; KANEMOTO, K.; FORAN, B.; LOBEFARO, L.; GESCHKE, A. *International trade drives biodiversity threats in developing nations*. *Nature*, v. 486, p. 109-112, 2012, p. 111.

LEVY, Steven. *The AI revolution is on.* WIRED, dez. 2010. Disponível em: <https://www.wired.com/2010/12/ff_ai_essay_airevolution/> Acesso em: 21 abr. 2017.

MILARÉ, Édis. *Direito do ambiente***.** 9ª ed. São Paulo: Editora Revista dos Tribunais, 2014.

RUNCIMAN, Brian. *Where are we with big data?* In: BCS, The Chartered Institute for IT. *Big data: opportunities and challenges.* Swindon: BCS Learning and Development Ltd., 2014.

RUSSEL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A modern approach***.** 3ª ed. New Jersey: Prentice Hall, 2010.

SKA - Square Kilometre Array. *Página Web*. Disponível em <https://www.skatelescope.org/>. Acesso em: 21 abr. 2017.

SURDEN, Harry. *Machine Learning and Law.* Washington Law Review, v. 89, n. 1, p. 87-115, 2014.

THE CORNELL LAB OF ORNITHOLOGY. *Página Web*. Disponível em: <http://www.birds.cornell.edu/page.aspx?pid=1609>. Acesso em: 30 abr. 2017.

THE IUCN RED LIST OF THREATENED SPECIES. *Página Web*. Disponível em: < http://www.iucnredlist.org/about/introduction>. Acesso em: 02 mai. 2017.

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY. ABOUT THE NATIONAL CENTER FOR COMPUTATIONAL TOXICOLOGY (NCCT). *Página Web*. Disponível em: <https://www.epa.gov/aboutepa/about-national-center-computational-toxicology-ncct>. Acesso em: 02 mai. 2017.

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY. EMISSIONS & GENERATION RESOURCE INTEGRATED DATABASe (eGRID). *Página Web*. Disponível em: <https://www.epa.gov/energy/emissions-generation-resource-integrated-database-e-

grid>. Acesso em: 02 mai. 2017.

WITTEN, Ian H.; FRANK, Eibe; HALL, Mark A. *Data mining: practical machine learning tools and techniques.* 3ª ed. Burlington: Elsevier, 2011.

Recebido em: 14/08/2017.

Artigo aceito em: 09/05/2018.

Como citar este artigo (ABNT):

MOLINARO, Carlos Alberto; LEAL, Augusto Antônio Fontanive. BIG DATA, MACHINE LEARNING E A PRESERVAÇÃO AMBIENTAL: INSTRUMENTOS TECNOLÓGICOS EM DEFESA DO MEIO AMBIENTE. *Veredas do Direito*, Belo Horizonte, v. 15, n. 31, p. 201-224, jan./abr. 2018. Disponível em: <http://www.domhelder.edu.br/revista/index.php/veredas/article/view/1142>. Acesso em: dia mês. ano.