
BIG DATA, MACHINE LEARNING Y LA PRESERVACIÓN AMBIENTAL: INSTRUMENTOS TECNOLÓGICOS EN DEFENSA DEL MEDIO AMBIENTE

Carlos Alberto Molinaro

Profesor en el PPGD De la Facultad De Derecho De la Pontificia Universidad Católica de Rio Grande Del Sul- PUCRS
Correo electrónico: carlos.molinaro@puers.br

Augusto Antônio Fontanive Leal

Doctorando en Derecho por la Pontificia Universidad Católica de Rio Grande del Sul (PUCRS). Master y graduado en Derecho por la Universidad de Caxias del Sur (UCS). Abogado.
Correo electrónico: augustoafl@gmail.com

RESUMEN

El presente artículo tiene por objetivo realizar un análisis explicativo sobre el contenido de las nuevas tecnologías denominadas de Big Data y *machine learning* y, con ello, efectuar una aproximación de la utilización de estas tecnologías como herramientas de protección ambiental. Para eso, la investigación se dividirá en dos tópicos esenciales: a) el enfoque y conceptualización de los recursos tecnológicos del Big Data y de las técnicas de *machine learning*; b) una exposición del perfeccionamiento derivado de la utilización del Big Data y de técnicas de machine learning para la finalidad de preservación ambiental. Por Intermedio de eso se podrá comprobar, a partir de ejemplos factibles, cómo la utilización de tecnologías como Big Data y machine learning pueden influir y demostrar un nuevo rumbo para la preservación ambiental, reforzando el deber del Estado y de la Sociedad en la protección y defensa de la naturaleza. La investigación será cualitativa amparada por el método hermenéutico con uso de investigación bibliográfica y documental.

Palabras-clave: Big Data; *machine learning*; preservación ambiental; medio ambiente.

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

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.*

INTRODUCCIÓN

Debido a los constantes impactos generados sobre la biodiversidad en un escenario global, ya es posible prever una calamidad ecológica *in fieri*. La complejidad de los problemas ambientales está enredada en cuestiones de difícil comprensión, especialmente en lo que toca a riesgos ambientales. Del mismo modo, las cuestiones relativas a los principios de la prevención y de la precaución también son acceso dificultoso ante las distintas dependencias encontradas en la biodiversidad, a partir de la estructura del medio ambiente y de las diversas relaciones y consecuencias desencadenadas por cada acto humano. No obstante, parece que, con la utilización de recursos tecnológicos, puede ser posible alcanzar medios más eficientes de combatir la degradación ambiental, no sólo a niveles locales, sino de modo global. En este sentido, la incontestable evolución exponencial de la tecnología permitió que actualmente se pudiera contar con las buenas prácticas del Big Data y del *machine learning*.

Es por medio de este estudio que se buscará establecer una relación entre innovaciones tecnológicas de gran porte, como es el caso del Big Data y de los algoritmos de *machine learning*, viabilizando una herramienta de protección ambiental con mayor precisión que sólo una gran cantidad de datos y una portentosa capacidad de procesamiento podría ofrecer. Para esto, el presente ensayo está dividido en dos capítulos. En el primero, se intentará un abordaje sobre la metódica del Big Data y las técnicas de *machine learning*, propiciando su comprensión. En el segundo, se buscará efectuar una aproximación de estas innovaciones tecnológicas en la propuesta de preservación ambiental, con el estudio de programas existentes que utilizan estas tecnologías para aumentar y mejorar la conservación de la naturaleza.

1. BIG DATA Y MACHINE LEARNING: ENFOQUE Y CONCEPTOS

En un mundo conectado, los datos se generan constantemente. Varios son los aparatos tecnológicos que posibilitan esta generación creciente de datos y que hace posible el establecimiento de patrones, una comprensión integral del mundo o, incluso, la capacidad de previsión de acontecimientos. Con el crecimiento exponencial tecnológico, la velocidad de procesadores y la capacidad de almacenamiento a gran escala hicieron posible la voluminosa figura del Big Data y el procesamiento de estos datos

posibilitando una comprensión que, sin técnicas de *machine learning*, no sería posible. Para disponer sobre estas temáticas, este capítulo se dividirá en dos secciones, siendo que en la primera se estudiará el Big Data y las oportunidades que son proporcionadas por su crecimiento incesante; en la segunda, se estudiará la revolución cognitiva a través de técnicas de *machine learning* de modo que se puedan encontrar patrones en la enorme cantidad de datos generados, permitiendo la comprensión de esa herramienta tecnológica en desarrollo.

1.1 Big Data: Mirada hacia un nuevo mundo de oportunidades

Por cierto, la utilización de datos ha aumentado por las sofisticaciones ofrecidas por la tecnología y sus incesantes avances, pero no se trata de una tarea derivada del surgimiento de ordenadores. Hace mucho tiempo se buscan patrones en datos recolectados, como es el caso de la búsqueda efectuada por un granjero por patrones en una cultura, o de un político en la opinión de un elector (WITTEN; FRANK; HALL, 2011, p. 4). Los datos siempre fueron almacenados de alguna forma en la vivencia humana, para las más diversas cuestiones, sea en un aspecto individual como en la memoria, sea en documentos y registros. Si en un determinado momento histórico el almacenamiento de datos era realizado por un cazador que buscaba comprender un patrón en el comportamiento migratorio de su caza, esto pudo haber pasado para el almacenamiento de datos para facilitar el comercio.

Sin embargo, en un entorno de desarrollo tecnológico, se puede decir que se está produciendo una sobrecarga de datos. Como abordan Witten, Frank y Hall, la cantidad de datos en el mundo y en la vida humana están creciendo y no hay un final para ello (WITTEN, FRANK, HALL, 2011, p.3). Según los autores, esto ocurre debido a la omnipresencia de computadoras en la vida cotidiana que, por medio de no tan caros sitios de almacenamiento, como es el caso de almacenamiento en línea, traen consigo la facilidad de salvar lo que anteriormente habría sido dispensado (WITTEN; FRANK; HALL, 2011, p. 4). Los datos simplemente se almacenan a lo largo de los días, por medios virtuales, en un escenario de facilidad tal que en cualquier momento puede ser la memoria de cualquier dispositivo ampliado para comportar un mayor número. Sin embargo, la generación de datos no se atiene a tan sólo la vida común de cada ser humano, sino que está en innumerables actividades corporativas,

gubernamentales, económicas y sociales.

Demostrando la capacidad de desarrollo de procesadores (CPU) todavía en 1965, Gordon Moore comprobó que el número de *transistors per square* doblaría cada año y, en el año 1975, revisando su teoría, Moore dobló este plazo para dos años (BELL, 2016). En consecuencia, hay una disminución del costo para el almacenamiento de datos (KATZ, 2013, p. 916), proporcionando un campo fértil para el desarrollo de lo que se acuñó de Big Data.

¿Cuán grande es el Big Data?

Katz remite a un objetivo subjetivo y siempre en movimiento, de modo que uno de los mejores caminos para comprender este concepto sería una mirada hacia atrás (KATZ, 2013, p. 916). La mejora continua de los sistemas de memoria, en conjunto con una capacidad de procesamiento se ha ampliado cada vez más. Eso demuestra que, en un *continuum*, hay una posibilidad de almacenamiento simplemente incapaz de ser comprendida en dispositivos electrónicos comunes y aislados.

Sin embargo, la gran fecha permanece como un concepto abstracto, de modo que, a pesar del amplio reconocimiento de su importancia, flotan diferentes opiniones sobre su definición, pudiendo ser conceptualizado, en general, como *datasets* que no pueden ser percibidos, adquiridos, administrados y procesados por tradicionales softwares y hardwares dentro de un tiempo tolerable (CHEN et al., 2014, p. 2). Según la definición traída por el International Data Corporation (IDC), Big Data describe una nueva generación de tecnologías y arquitecturas, diseñada para económicamente extraer valores de inmensos volúmenes de una amplia variedad de datos, a través de una muy veloz captura, descubrimiento y/o análisis (GANTZ; REINSEL, 2011, p. 6). Por lo tanto, es importante aclarar que Big Data no trata solamente del contenido original almacenado o consumido, sino también de la información envuelta de este consumo, lo que puede ser ejemplificado por la forma en que los smartphones producen una fuente de datos adicionales que se están capturando, como es el caso de la ubicación geográfica, mensajes de texto e historial de navegadores de Internet (GANTZ; REINSEL, 2011, p. 6).

Generados por los más diversos medios tecnológicos, los datos en una cantidad absurdamente difícil de ser notada, aprehendida y mensurada no atienden a un crecimiento lineal, sino exponencial. Como explicó Katz, hay un desarrollo exponencial en una sinergia entre los conceptos de Ley de Moore, Big Data, y la revolución de la inteligencia artificial, de modo

que cada duplicación de la velocidad de los procesadores, de la división a la mitad del precio para el almacenamiento de datos y avances del *machine learning*, abren una inmensidad de posibilidades en una perspectiva no lineal (KATZ, 2013, p. 922).

En este escenario horizontalmente transversal, Big Data se forma a partir de diversas fuentes, pudiendo incluir variados tipos de datos en una cantidad difícil de ser absorbida. En este gran número, las nuevas fuentes de datos presentes en los medios sociales en los que los usuarios, a veces incluso inconscientemente, acaban generando un gran volumen de datos que pueden ser utilizado en variadas formas y que se van expandiendo en rápidas tasas algorítmicas. Cómo describen Gantz y Reinsel (2011, p. 6):

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. (Los grandes datos son una sección transversal horizontal del universo digital y pueden incluir datos transaccionales, datos almacenados, metadatos y otros datos que están en archivos especialmente grandes. Medios de comunicación/entretenimiento, cuidados de salud y vigilancia por video son ejemplos claros de nuevos segmentos de gran crecimiento de datos. Soluciones de redes sociales como Facebook, Foursquare y Twitter son las nuevas fuentes de datos más recientes. Esencialmente, ellos crearon sistemas donde los consumidores (conscientemente o inconscientemente) están forneciendo flujos de datos casi continuos sobre sí mismos, y gracias al “efecto de red” de sitios bien-sucedidos, los datos generados pueden expandir en tasas logarítmicas rápidas - Trad. Libre.)

Sin embargo, aunque los medios sociales representen una gran contribución en la generación de datos, no se trata de la única fuente. Además de las plataformas de medios sociales que producen una gran cantidad de datos por las redes de perfiles individuales o por otros contenidos producidos, que tratan de blogs, foros de discusión de usuarios, incluso “gustar”, contiene potenciales informaciones a ser usadas, pudiendo ser se mencionan como fuentes los datos producidos por actividades como

el contenido mecánicamente generado en forma de archivos de registro de dispositivos, en la función de internet de las cosas (*internet of things*). Otras fuentes de datos pueden todavía encontrarse en servicios de software, aplicaciones en nube y las hasta entonces un tanto inexploradas, son los datos de registros médicos y correspondencia con clientes (RUNCIMAN, 2014, p.1). No es casualidad que, según IBM (2017), diariamente se crean 2,5 quintiles de bytes de datos. Es el 90% de los datos en el mundo hoy en día se crearon en los últimos dos años, tratándose de datos que vinieron de todo lugar: sensores de información sobre el clima, posturas en medios sociales, fotos digitales, videos, registros transaccionales, registros de transacciones señales de GPS de teléfonos celulares, entre muchas otras formas.

Las posibilidades existentes a partir de la utilización del Big Data son las más diversas y representan un fértil campo para perfeccionamiento de la vida humana y de todas las especies de vida de la tierra. Es necesario disponer de medios para la utilización de estos datos que contemplen su identificación, verificación, comprensión y utilización. Basu y Hall, demostrando la capacidad de la utilización de este recurso tecnológico, citan la presencia del Big Data en el *Square Kilometre Array*, un radio telescopio con tamaño capaz de hacerlo cincuenta veces más sensitivo y diez mil veces más rápido que cualquier otro, que irá generar por día la cantidad de datos existentes en Internet, haciendo astrónomos capaces de percibir cualquier señal en la galaxia y mirar más profundamente por el espacio como nunca antes (BASU; HALL, 2014, p. 46); de acuerdo con esos mismos autores, el Big Data es parte de una revolución en el mundo doméstico, social y de los negocios, pues así como el macizo radio telescópico *Square Kilometre Array* va a generar Big Data para habilitar un mejor entendimiento de la historia y lugar en el espacio del hombre, existen grandes oportunidades para incrementar el uso de Big Data para entender mejor el futuro ambiente humano y las posibilidades en el planeta (p. 51).

En consecuencia de un desarrollo tecnológico exponencial, con procesadores cada vez más potentes y la disminución del costo de almacenamiento de datos, se posibilitó una generación de datos de gigantesca cantidad y que crece diariamente, teniendo como fuente la creación de datos, los conscientes o inconscientes usuarios de medios sociales o demás dispositivos conectados, sin contar el registro realizado mecánicamente de diversos aparatos con conexión (internet de las cosas), el

almacenamiento en nube y diversos otros datos que pueden ser explotados. Estos datos a gran escala que ayudan a construir el significado de Big Data ofrecen innumerables posibilidades y oportunidades para revolucionar positivamente la vida humana en la Tierra y, como es en el caso de la problemática ambiental, propiciando un campo fértil en la protección ambiental.

Sin embargo, no basta que se reconozca la existencia del Big Data. Son necesarios medios para que estos datos puedan ser utilizados, es decir, hay que descubrir o canalizar la gran cantidad de datos por medios que hagan posible su utilización. Como se ve, se trata de una cantidad de datos absurdamente enorme y no hay forma en que el intelecto humano para hacer de este servicio sin una herramienta tecnológica capaz de realizar la minería de estos datos, seleccionar lo que se busca y verificar, a través de un análisis, lo que se quiere. Es en este sentido que surge el *machine learning* y una revolución en el campo de la inteligencia artificial, como se demostrará en el próximo subcapítulo.

1.2 Revolución cognitiva, *machine learning* y la minería de datos

El colosal volumen de datos que día tras día es generado para ser utilizado de acuerdo con una propuesta de alcanzar un perfeccionamiento en la vivencia humana, por las oportunidades que existen y vendrán a existir con el Big Data, depende de un aparato tecnológico que dé cuenta, no sólo de procesar e interpretar de modo hercúleo una determinada cantidad de datos, pero que realice tareas de elección y pueda, efectivamente, aprender con el análisis generado para posibilitar respuestas de análisis de comportamiento, identificando patrones que puedan ser utilizados.

Aunque una herramienta de este porte sea dirigida por algún ser humano, es imposible concebir que la mente humana pueda trabajar con tamaño volumen de datos como se demuestra en lo que se acuñó de Big Data. Como evidenciado por Witten, Frank y Hall, hay una gran brecha entre la generación de datos y el entendimiento humano de estos datos (WITTEN; FRANK; HALL, 2011, p. 4). Parece que, mientras el volumen de datos sigue aumentando, la aptitud humana para comprenderlo disminuye de manera alarmante, haciendo que una gran cantidad de datos permanezca no utilizada (WITTEN; FRANK; HALL, 2011, p. 4). Estos datos, con potencial utilidad, raramente vienen a ser identificados o utilizados, permaneciendo en un estado de invisibilidad. La minería de

datos (*data mining*) no es un recurso reciente, ya que los economistas o los ingenieros de comunicación, por ejemplo, hace mucho trabajan con la idea de encontrar patrones en datos automáticamente, identificando, validando y utilizando para previsiones. Lo que hay de nuevo es un incremento asombroso en las oportunidades para que se encuentren estos patrones en datos que provienen del insaciable crecimiento (WITTEN; FRANK; HALL, 2011, p. 4).

Se puede notar que, según Katz, la capacidad de almacenamiento de datos y la velocidad de procesadores no son suficientes, por sí solos, de generar el conjunto de percepciones agregadas que conduce a una mayor productividad e innovación. Según el autor, además de estas dos innovaciones tecnológicas, una revolución de la inteligencia artificial debe ser observada como un tercer pilar en una nueva era de productividad (KATZ, 2013, p. 918). Es en este contexto que el uso de técnicas de *machine learning* (que en este texto se utilice el término original que puede ser entendido en portugués como aprendizaje computacional o aprendizaje automático) que aparece como una herramienta para encontrar y describir patrones en los datos (WITTEN; FRANK; HALL, 2011, p. 5).

Sin embargo, hay un contenido metafórico en decir que un ordenador puede aprender, aparentando una expresión meramente ilustrativa. Analizando cinco definiciones sobre aprender: a) tener conocimiento de algo por el estudio, experiencia, o aprendiendo; b) hacerse consciente de una observación o información; c) memorizar; d) ser informado o averiguar algo; e) recibir instrucción, preguntan Witten, Frank y Hall sobre cómo se podría saber si una computadora ha aprendido algo? ¿O si un ordenador se ha enterado de algo? La cuestión que se refiere a la posibilidad de que un ordenador esté informado o consciente es una cuestión actual de la filosofía. No obstante, los últimos tres aspectos (ítems “c”, “d” y “e”), podrían ser muy aproximados al significado de *machine learning* (WITTEN, FRANK, HALL, 2011, p. 7). En el caso de que, en vez de buscar una réplica del proceso cognitivo humano en computadoras que estaba basada en la creencia de crear artificiosas versiones de la funcionalidad cerebral humana, quedando adscrito a un supuesto proceso de producción inteligente, está la posibilidad alternativa de obtener resultados que los procesos y en el caso de que se trate de un proceso cognitivo humano (SURDEN, 2014, p. 98-99).

De este modo, el aprendizaje se podría atribuir a las cosas cuando estas cambian su comportamiento de una manera que puedan tener un

mejor desempeño en el futuro, tratándose mucho más de una cuestión de desempeño que de conocimiento (WITTEN; FRANK; HALL, 2011, p. 7). Como se mencionó por Surden, funcionando bien, los algoritmos de *machine learning* pueden producir resultados automáticos aproximados de lo que realizarían una persona igualmente situada (SURDEN, 2014, p. 90). Así, la utilización de datos, por la minería y posterior verificación con el fin de encontrar un patrón de comportamiento que pueda ser utilizado no sólo para comprender algo ocurrido o que esté ocurriendo, sino para disponer sobre una previsión, ante una gran cantidad de datos (Big Data), se daría por técnicas de *machine learning*.

Con el ejemplo de la clasificación automática de e-mails como spam, Surden busca demostrar que los algoritmos de *machine learning* pueden desarrollarse cada vez más así recibir una mayor cantidad de datos, teniendo por meta la construcción de un modelo de un complejo fenómeno que permitirá un equipo de realizar clasificaciones precisas (SURDEN, 2014, p. 92). Con base en eso, se puede decir que computadoras pueden tener resultados “inteligentes” en tareas sin la intervención de la cognición humana (SURDEN, 2014, p. 95).

En suma, por el proceso de aprendizaje se tendría el perfeccionamiento de un comportamiento por el estudio de experiencias (RUSSEL; NORVIG, 2010, p. 693). Las técnicas de máquina de aprendizaje operan aprendiendo a partir de la experiencia y desarrollando una mejora de su desempeño por el tiempo, lo que contribuiría sobremedida en la medida en que fuera implementada en los más diversos sectores, como económicos, sociales, ambientales, capacitando una comprensión y propiciando previsiones.

Hay que decir, sin embargo, en un aspecto problemático presente en las previsiones. Si el objetivo de la utilización de técnicas de *machine learning* está pautado en el análisis de datos para obtener resultados aproximados en el futuro, en escenarios aún no vistos (o nunca vistos), se puede chocar con el problema de la generalización (SURDEN, 2014, p.105). Surden reparte el problema de la generalización en tres aspectos: a) si los acontecimientos futuros dependan de hechos únicos o no inusual, la previsión acabaría dificultada; b) se debe evitar una sobregeneralización (overgeneralization), pues es indeseable que los algoritmos de *machine learning* detecten datos idiosincrásicos o prejuiciosos que, obviamente, no tendrían condiciones de hacer una previsión; c) es el caso, por ejemplo, de previsiones legales, cuando algún algoritmo de *machine learning* encuentre

dificultades de capturar sutiles factores de gran relevancia (SURDEN, 2014, p. 105-106).

El problema de la generalización podría hacer caer por tierra una previsión que se buscó encontrar cuando algún factor clave no esté presente en los datos anteriores, cuando exista algún caso en que los datos están dotados de falsas percepciones como, por ejemplo, pre-juicios que impedirían una el análisis correspondiente con la realidad y, aún, cuando existieran factores sutiles que difícilmente serían captados por la técnica de *machine learning* utilizada. Para actuar sobre estos factores, es importante que los datos utilizados sean filtrados en un criterio tal que se pueda medir ciertas características. También, ciertos problemas existentes en la posible ocurrencia de atipicidades o de factores de difícil percepción pueden ser perfeccionados a lo largo del uso del sistema de *machine learning* desenvuelto.

Por eso, para maximizar la potencia de técnicas de machine learning, en una teoría de aprendizaje computacional que esté pautada en la intersección entre inteligencia artificial, estadística y teoría de la ciencia computacional, se demuestra un principio subyacente de que cualquier hipótesis que sea seriamente equivocada casi sería encontrada, tras un pequeño número de ejemplos, pues estaría haciendo una predicción errónea, de modo que cualquier hipótesis consistente con un amplio sistema de ejemplos demostraría un improbable error, estando probablemente aproximadamente correcta (*probably approximately correct*) (RUSSEL; NORVIG, 2010, p. 714). Para eso, la capacidad de procesar una gran cantidad de datos, probando hipótesis con una aproximación cada vez mayor de una respuesta correcta y coherente, tendría la función de eliminar problemas derivados de una generalización, incluso cuando para una previsión se dependiera de un hecho que aún no ocurrió, pues una serie de experiencias llevaría, después de un determinado período, en la descaracterización de este problema por el contenido de circunstancias hasta entonces analizadas. Concluyendo, Flach concibe *machine learning* como el estudio sistemático de algoritmos y sistemas con la finalidad de mejorar su conocimiento o desempeño por la experiencia (FLACH, 2012, p.3). A lo largo del tiempo, las técnicas de *machine learning* se vuelven más provechosas y permiten, ante una gran cantidad de datos, un perfeccionamiento de diversos sectores de la sociedad, revolucionando la vida humana.

La utilización de *machine learning* que, a través de imágenes generadas diariamente por satélites radares que monitorean aguas costeras

para detectar manchas de aceite, permite entrenar un sistema detector de contaminación, con ejemplos sobre derrames para cada usuario mediante sus propios propósitos es sólo un ejemplo sobre como esta tecnología puede ayudar en la preservación ambiental. En el caso, al encontrar posibles manchas de aceite, primero el sistema procesa y normaliza la imagen, identificando la mancha sospechosa. Después, se determinan los atributos de cada mancha de aceite, como la región, el tamaño y la intensidad y, al final, los aprendizajes de patrones se utilizan para los atributos resultantes (WITTEN; FRANK; HALL, 2011, p. 23).

Como se afirma en Flach, existen grandes posibilidades de que los algoritmos de *machine learning* ya estén siendo utilizados por un alto número de personas, aunque éstas no lo sepan (FLACH, 2012, p.1). De igual forma, Russel y Norvig refieren la existencia de una actuación detrás de las escenas de la inteligencia artificial, como en el caso de la automática aprobación de compra de tarjeta de crédito. Estos cambios que se incorporan acaban modificando la vida diaria de las personas, aunque muchas no se den cuenta de ello. A pesar de estar en desarrollo, se puede esperar que un mediano éxito de la inteligencia artificial vendría a afectar a todas las personas en sus vidas (RUSSEL; NORVIG, 2010, p. 1051).

Lo cierto es que el mundo como es conocido viene cambiando diariamente con el exponencial avance tecnológico de procesadores, técnicas de almacenamiento de memoria e incremento de algoritmos de *machine learning*. Se ofrece así un vasto campo de posibilidades y oportunidades para mejorar la vida humana en diversos aspectos, entre los cuales y seguramente uno de los más importantes, está la preservación de la naturaleza como un todo. Incluso, algunas técnicas ya se están desarrollando y sirven como ejemplos sobre cómo la tecnología de Big Data y de *machine learning* pueden traer efectividad para asegurar un medio ambiente ecológicamente equilibrado, tanto a escala local como global. Es lo que se demostrará en el próximo capítulo.

2. LA FUNCIÓN DE LA TECNOLOGÍA DEL BIG DATA Y LAS TÉCNICAS DE *MACHINE LEARNING* EN LA PRESERVACIÓN AMBIENTAL

La utilización de algoritmos de *machine learning* aliados a una capacidad portentosa de almacenamiento de datos representan conjuntamente en significativos cambios en los más variados aspectos

de la vida humana. Por eso, tanto Katz (2013) cuanto Surden (2014) buscan relacionar el uso de *machine learning* para el mejor desempeño de tareas legales, el primer autor, trabaja en una nueva era de previsión legal cuantitativa que vendría a auxiliar abogados, estudiantes y facultades de Derecho (p. 912-913); ya el segundo, advierte que no es insignificante, número de tareas legales que pueden ser parcialmente automatizadas usando técnicas de inteligencia artificial (p. 88).

En la búsqueda por la preservación de la naturaleza eso no es diferente (!).

El desarrollo tecnológico amplía el instrumental que sirve tanto para encontrar los problemas ambientales como para alcanzar soluciones. Tal perspectiva puede traer una mayor efectividad al derecho fundamental al medio ambiente y sus determinaciones previstas en el artículo 225 de la Constitución Federal brasileña (BRASIL, 1988), como es el caso de la defensa y preservación de un medio ambiente ecológicamente equilibrado, teniendo en vista la realidad de un asunto *res maximi momenti* que llevó a estar constitucionalizados los fundamentos de protección ambiental y del incremento de su calidad (MILARÉ, 2014, p. 160-161). Sin embargo, esta preocupación con la realidad natural no está asociada a una mera positivación. Además, el medio ambiente alcanza los niveles de un derecho fundamental que conduce a la formulación de un principio de primacía del ambiente, reconociendo la ilicitud de tratar el medio ambiente como subsidiario, accesorio, menor o despreciable (BENJAMIN, 2010, p. 118).

Se considera que la propuesta de un proyecto como Global Forest Watch (<http://www.globalforestwatch.org/>), que consiste en un sistema de alerta y monitoreo de bosques online que ofrece herramientas para que cualquier persona del mundo pueda ser capaz de gestionar y conservar las florestas, es un claro ejemplo de la utilización de herramientas tecnológicas para la preservación ambiental. El triunfo tecnológico de un proyecto de esta magnitud no está solamente pautado en una gigantesca recolección de datos de nivel global (Big Data), sino también en una unión de personas de todo el mundo en el suministro de datos y el uso de sistemas de alerta que viabilizan una mayor protección forestal.

El anteriormente citado artículo 225 de la Constitución Federal impone, tanto al Estado como a la colectividad, la preservación y defensa del medio ambiente ecológicamente equilibrado. Justamente por eso, según Sarlet y Fensterseifer, subsiste una diferencia entre este derecho y los derechos fundamentales de primera dimensión, pautados en los derechos

liberales, así como del principio de igualdad, asentado en derechos sociales de segunda dimensión, concluyendo por el pilar normativo-axiológico del derecho fundamental al medio ambiente como principio de deber de solidaridad (SARLET; FENSTERSEIFER, 2013, p. 47). Con eso, el propio sistema cooperativista en proyectos de alta punta tecnológica, utilizando Big Data, como el Global Florest Watch posibilitan medios para que los ciudadanos y el propio gobierno brasileño tengan condiciones de cumplir con su deber de protección ambiental presente en el contenido normativo del artículo 225 de la Constitución Federal. Además de las regulaciones internas, se debe tener en cuenta también el cumplimiento de tratados internacionales que se refieren a las más diversas cuestiones de nivel global, demostrando que no sólo para una ejecución eficaz de normas internacionales, el uso de Big Data y *machine learning* sirven para establecer tratados justos y que tengan un contenido preciso y necesario de reglamentación.

Más profundamente, sin embargo, se buscará demostrar en los siguientes subcapítulos como significativos impactos tecnológicos como Big Data y técnicas de *machine learning* pueden influenciar y reorientar los esfuerzos en la preservación ambiental a través de un perfeccionamiento general de las herramientas a ser utilizadas.

2.1 BirdCast y la migración de pájaros en tiempo real

BirdCast (<http://birdcast.info/>) es un programa desarrollado con la finalidad de, por primera vez, viabilizar previsiones de migraciones de pájaros, estableciendo condiciones para saber cuándo ellos van a migrar, a donde ellos migrarán y cuánto volarán. El proyecto, desarrollado por The Cornell Lab of Ornithology, propone que el conocimiento del comportamiento migratorio auxiliará en criterios de conservación, como es el caso del posicionamiento de turbinas eólicas o incluso en la identificación de que en determinada noche algún alto edificio podrá tener que dejar sus propias luces encendidas previniendo la muerte de millones de pájaros. Sin embargo, la importancia de BirdCast va más allá. Higuchi, al mencionar haber colaborado con científicos estadounidenses y asiáticos en el rastreo por satélite de la migración de aves por veinte años, demuestra la importancia del conocimiento migratorio de pájaros para la conservación del medio ambiente global (HIGUCHI, 2012, p.3). El autor señala que los pájaros migratorios encuentran varios problemas durante la migración,

lo que incluye la destrucción de hábitats, contaminación química, caza y colisiones con aviones (HIGUCHI, 2012, p. 3).

Hay una conexión entre pájaros migratorios y cuestiones futuras sobre la relación el medio ambiente y las actividades humanas. Cada otoño y primavera, un gran número de aves visitan diversas áreas en sus rutas migratorias y, como sin comida no hay migración, obtienen comida y otras cuestiones significativas en estas áreas. En consecuencia, los pájaros migratorios tienen un papel importante en el mantenimiento de ecosistemas en diferentes áreas como consecuencia de sus hábitos alimenticios, de modo que, si sólo existen pocos pájaros migratorios, la población de insectos podría aumentar drásticamente en los bosques causando un desequilibrio en la naturaleza, así como algas podrían crecer densamente en los lagos, contaminando el agua (HIGUCHI, 2012, p. 11). A veces, estos impactos negativos resultantes de un cambio en la migración de pájaros no son bien comprendidos. La disminución de la población de aves migratorias en un determinado hábitat puede acarrear el deterioro de un ecosistema en otra área. Según Higuchi, la destrucción de bosques tropicales en el sudeste asiático disminuye el número de pájaros migratorios tropicales que allí pasan el invierno y emigran a Japón en la primavera, generando un desequilibrio en la naturaleza de los ecosistemas forestales en este país. También, la destrucción de planicies de marea en Japón puede disminuir el número de muchas aves limícolas, provocando un deterioro de los ecosistemas de zonas húmedas en Filipinas, Australia y Rusia (HIGUCHI, 2012, p. 12). Por eso, cuestiones como éstas representan un desafío para el desarrollo de la ornitología y la ecología, evidenciando que la conservación de pájaros no se vuelve sólo para su preservación, sino también para el mantenimiento de ecosistemas, llevando a la conservación del medio ambiente global (HIGUCHI, 2012, p. 12).

El programa BirdCast ofrece modelos precisos de migración de pájaros, permitiendo a los investigadores comprender los aspectos de la migración, como el tiempo de migración y las vías de migración responden a los cambios climáticos, y cuales conexiones existen entre el tiempo de migración y los cambios de tamaño poblacional. Entre las fuentes de datos del programa están las llamadas de vuelo, que permiten identificar la composición de especies de migración nocturna, cuando de su ocurrencia, proporcionando informaciones en tiempo casi real, los radares de vigilancia meteorológica, que pueden ser utilizados para observar y rastrear los movimientos de los pájaros y, por último, el revolucionario

eBird.

Iniciado en 2002, por The Cornell Lab of Ornithology y el National Audubon Society, eBird es un programa que ofrece una lista en tiempo real para comunicar informes y acceder a información sobre los pájaros, proporcionando datos de fuentes para información básica en la abundancia de pájaros y la distribución en escalas espaciales y temporales. El programa eBird es uno de los mayores y más rápidos acumuladores de recursos de datos de biodiversidad que existen. En mayo de 2015, más de nueve millones de informes de observaciones de aves a través del mundo fueron recolectados por el programa. Basado en estas fuentes de datos, incluso con el portentoso recurso de eBird, el programa BirdCast propone el desarrollo de dos técnicas de *machine learning*, Collective Graphical Models (CGMs) y Semi-Parametric Latent Process Models (SLPMs). Las dos técnicas tienen por objetivo, conjuntamente, reconstruir y prever el comportamiento de cuatrocientas especies de pájaros migratorios en América del Norte, resultando en un modelo capaz de identificar las complejas condiciones que gobiernan la dinámica del comportamiento migratorio. Además, BirdCast busca el desarrollo de datos interoperables para sintetizar observaciones de pájaros, llamadas de vuelo, datos de radares y datos de múltiples fuentes, como imágenes de satélite, meteorología y datos de poblaciones humanas. Contando, al fin y al cabo, con un sistema de datos en web para comunicar predicciones migratorias generadas por el BirdCast al público.

Con la utilización de técnicas de *machine learning* junto a una inmensa fuente de datos acumulativos, BirdCast permite la comprensión del comportamiento migratorio de pájaros, proporcionando valiosas informaciones para aumentar significativamente la preservación ambiental a escala global.

2.2 Cadena logística global y la identificación de puntos centrales de especies amenazadas

La cadena logística representa la trayectoria de un producto, desde concepción como un suministro, pasando por la etapa de producción en un proceso de industrialización y alcanzando, al final, la distribución, que atiende a la demanda de los consumidores. No es raro, este proceso se da en un nivel globalizado, generando una interconexión de diversas partes en este complejo proceso de producción. Las compañías que actúan sobre

áreas críticas para especies amenazadas, sin embargo, tienden a continuar con la sobreexplotación agravando aún más la problemática retratada.

La pérdida de la biodiversidad y, más precisamente, las amenazas locales a las especies están siendo conducidas por la actividad económica y la demanda de consumo (LENZEN et al., 2012, p.111). Por lo tanto, el 30% de las especies de todo el mundo están amenazadas por el comercio internacional (LENZEN et al., 2012, p. 109). Comprender esto demuestra que no se tiene como actuar solamente a nivel local, en torno a una regulación que no actúe solamente sobre países en desarrollo que realizan exportaciones en alto número, existiendo la necesidad de que regulaciones alcancen países donde el consumo actúa activamente en la creación de una gravosa huella ambiental que pone en peligro innumerables especies.

La cuestión está justamente en el escenario que implica la identificación de puntos centrales de especies amenazadas a partir de una cadena logística global. Y es justamente este el triunfo de una investigación realizada pautada en la construcción de un mapa que combina diversos mapas de ocurrencia de una serie de especies amenazadas. Con estos datos espaciales de una mirada ambiental, es posible saber qué países y qué categorías de consumo amenazan hábitats en varios puntos críticos (MORAN; KANEMOTO, 2017, p. 1). La investigación trae resultados notables, como es el caso de la mirada ambiental sobre especies terrestres proveniente de los Estados Unidos, no sólo sobre el sudeste asiático o Madagascar, como se esperaba, pero también en puntos críticos al sur de Europa, en el Sahel, en la costa este y oeste del sur de México, Centroamérica y Asia Central y en el sur de Canadá. A pesar de haber una atención redoblada sobre la selva amazónica, el estudio reveló que la huella de Estados Unidos en Brasil es mayor en la meseta, al sur, que en la Amazonia. La profunda huella de los Estados Unidos en el sur de la biodiversidad de España y Portugal, en relación con el impacto en el número de especies de aves y peces amenazados, es digno de mención, ya que estas áreas rara vez son consideradas como como punto crítico en lo que se refiere a especies amenazadas (MORAN; KANEMOTO, 2017, p. 1). Por la investigación también fue posible observar puntos críticos de amenazas para varios países consumidores mayoritarios y acercarse a cada particular región afectada por el consumo. Es el caso del impacto del consumo de Estados Unidos en Sudáfrica, del consumo de Europa en África, y del sudeste de Asia por el consumo japonés (MORAN; KANEMOTO, 2017, p. 1).

Por la utilización del método de puntos críticos, se ha podido identificar qué áreas en que la amenaza a la biodiversidad es predominantemente conducida por un pequeño número de países. Con eso, identificando las regiones en que determinados países están presionando la biodiversidad, se permite el inicio de una directa colaboración entre productores y consumidores, atendiendo a las regulaciones internacionales, a fin de mitigar los impactos ambientales en los locales identificados (MORAN; KANEMOTO, 2017, p. 3). Contabilizando espacialmente los impactos explícitos en la biodiversidad de determinadas áreas, es posible ayudar a mejorar la producción sostenible, el comercio internacional y el consumo. La responsabilidad por los impactos ambientales debe repartirse a lo largo de toda la cadena de logística, de las industrias a los consumidores finales. Siendo que los mapas proporcionados por la investigación ayudan desde los productores a los consumidores, incluidos los conservacionistas, en la búsqueda de soluciones en el rastreo de puntos críticos para la biodiversidad (MORAN; KANEMOTO, 2017, p. 3).

La investigación realizada sólo se hizo posible por medio de un masivo contenido de datos de diversas fuentes. Los resultados se calcularon a partir de los tres datos disponibles (MORAN; KANEMOTO, 2017, p. 5):

1. IUCN Red List of Threatened Species: contiene la diversidad genética y los bloques de construcción de los ecosistemas, informaciones sobre los estados de conservación y distribución y capacitan decisiones informadas sobre la conservación de la biodiversidad a nivel local y global. Proporciona información taxonómica, estado de conservación y distribución acerca de plantas, hongos y animales evaluados globalmente.

2. BirdLife International: se trata de una asociación de nivel global, formada por diversos socios, siendo cada una, organización ambiental no gubernamental, sin fines de lucro. BirdLife International busca a través de una rigurosa ciencia, implementar programas de conservación para pájaros y toda la naturaleza. La asociación global permite un trabajo para la naturaleza en cada lugar en que hay un miembro establecido, estando éste conectado nacional e internacionalmente.

3. Eora MRIO: posee una base de datos de input-outputs con la ayuda de satélites sociales y ambientales en 187 países en 15.909 sectores.

La importancia de esta investigación que se ha valido de gigantescos bancos de datos demuestra cómo es necesario un acercamiento directo de la legislación de nivel internacional y local para proporcionar una más precisa preservación del medio ambiente en su completitud. Identificar

la cadena logística y las amenazas a la biodiversidad que ocurren en una determinada localidad pueden llevar a concebir tratos entre los diversos países involucrados, buscando una responsabilización compartida y una mayor conservación de la biodiversidad.

2.3 Programas gubernamentales tecnológicos: un ejemplo a partir de herramientas tecnológicas utilizadas por la Agencia de Protección Ambiental norteamericana

De acuerdo con Breggin y Amsalem, Big Data viene siendo utilizado por agencias gubernamentales, organizaciones no gubernamentales, y empresas privadas para proteger el medio ambiente, siendo que las autoras ejemplifican por medio de la mejora en la eficiencia energética, la promoción de la justicia ambiental, seguimiento del cambio climático y del monitoreo de la calidad del agua (BREGGIN; AMSALEM, 2014, p. 3). No es casualidad que, según constatan las autoras, en el gobierno federal estadounidense, la mayoría de las entidades gubernamentales responsables de la protección ambiental han usado Big Data en su trabajo. La utilización puede ser ilustrada por el análisis de riesgo, investigación, ejecución, educación pública y capacitación.

En la US Environmental Protection Agency (Agencia de Protección Ambiental de los Estados Unidos), se puede citar el monitoreo de calidad de aire que viene siendo elaborado en la oficina de investigación y desarrollo, teniendo como uno de los ejemplos el enfoque en nuevas tecnologías que auxilian al gobierno a evaluar la capa de ozono y la conformidad con los estándares de calidad del aire nacional junto con un método de fusión de datos que combine el resultado de los monitoreos para crear previsiones (BREGGIN; AMSALEM, 2014, p. 6).

Otro ejemplo es el desarrollo del CompTox, un programa de investigación en toxicología computacional, que tiene por propósito tratar con la ausencia de datos sobre salud y medio ambiente en miles de productos químicos (BREGGIN, AMSALEM, 2014, p.7). El programa utiliza una avanzada investigación de integración entre biología, biotecnología, productos químicos y ciencia de la computación para identificar importantes procesos biológicos que pueden desencadenarse de productos químicos, de modo que miles de productos químicos pueden tener sus potenciales riesgos evaluados en un corto espacio de tiempo.

Otro ejemplo es el Emissions & Generation Resource Integrated

Database (eGRID), un inventario de datos federales sobre plantas de energía y empresas de energía recolectadas, que abarca las características ambientales de casi toda la electricidad generada en los Estados Unidos. Entre las características están la emisión atmosférica de contaminantes, la tasa de emisión y la cantidad de energía eléctrica generada en una central eléctrica que es transmitida y distribuida al consumidor. Estas enormes cantidades de datos se agrupan por instalaciones, empresas y el Estado (BREGGIN; AMSALEM, 2014, p. 7).

Con el objetivo de unificar sistemas, combinando y conectando satélites y sistemas de observación ubicados en el suelo alrededor del mundo, el Global Earth Observation System of Systems (GEOSS) es un proyecto en desarrollo por The Group on Earth Observations (GEO) asociación internacional entre gobiernos y organizaciones. Se intenta ayudar a la comunidad internacional a coordinar respuestas emergentes a los desastres naturales o causados por el hombre, tales como incendios forestales, monitorear cambios climáticos y en la gestión de recursos naturales (BREGGIN; AMSALEM, 2014, p. 8). Por GEOOS, hay un aumento del entendimiento de los procesos de la Tierra y una mejora de las capacidades predictivas que sustentan la toma de decisiones, proporcionando acceso a datos, informaciones y conocimiento para una amplia variedad de usuarios.

La aproximación tecnológica entrante por la Agencia de Protección Ambiental demuestra la imprescindibilidad de la utilización de recursos tecnológicos de alta punta para una eficaz conservación del medio ambiente. La complejidad de los diversos factores que se refieren a la biodiversidad puede ser entendida por recursos computacionales que tengan la capacidad de almacenar una gran cantidad de datos, así como de procesar los datos recopilados proporcionando entendimiento para decisiones a tomar en el ámbito gubernamental. Además, la reglamentación ambiental puede tener mayor precisión por el análisis de una gran cantidad de datos, Big Data, en conjunto con el procesamiento de estos datos que pueden realizar previsiones sobre innumerables cuestiones, haciendo posible una anticipación con la finalidad de prevenir la ocurrencia de desastres ambientales, entre otros factores que afecten la biodiversidad.

CONCLUSIÓN

La revolución proporcionada por la capacidad de procesadores de computadora junto con el considerable aumento de los recursos de almacenamiento de datos llevó a la actual era del Big Data. La velocidad en la creación de datos en una magnífica cantidad que hace imposible su análisis por sistemas de gestión tradicionales ha aumentado las posibilidades para una mayor comprensión de diversos factores del mundo, como es el caso de la economía, la sociedad y la naturaleza. En conjunto con el Big Data, el advenimiento de técnicas de *machine learning*, posibilitando un procesamiento de esta gran cantidad de datos capacita esta comprensión, proporcionando un instrumento tecnológico como nunca antes visto y trayendo en los desafíos del desarrollo y perfeccionamiento de esta tecnología, posibilidades innumerables para los usuarios más variados sectores de la sociedad.

En el presente ensayo, se ha buscado aproximar la era del Big Data como forma de proteger y preservar el medio ambiente. La interconexión de la biodiversidad en diversas regiones demuestra el carácter difuso del bien ambiental y la comprensión de sus redes de conexiones a través de una gran cantidad de datos sobrepasa las barreras locales y permite una aproximación de los hechos como ocurren en el plan global estatal, regional y local. Con ello, se permite un justo diálogo entre las partes involucradas para establecer una equivalente responsabilización por la preservación ambiental, así como trae materias a ser reguladas en el plan internacional e interno de cada país.

Los proyectos existentes, sean los no gubernamentales formados por miembros de varios lugares del mundo en una cooperación por la preservación ambiental, sean los gobiernos, como es ejemplificado por los desarrollos de la Agencia de Protección Ambiental norteamericana, demuestran cómo el Big Data ha ayudado al gobierno en preservación de la naturaleza, alcanzando el puesto de pieza clave para decisiones más precisas. Concluyendo que la utilización de Big Data y de técnicas de *machine learning*, además de ser posible, ya es un sistema utilizado para preservar el medio ambiente, se puede afirmar que en el plan interno brasileño, la utilización de estos recursos tecnológicos debe ser implementada con la finalidad de traer eficacia al derecho fundamental al medio ambiente ecológicamente equilibrado, en las tomas de decisiones políticas, en las políticas públicas de justicia ambiental, en el ejercicio

jurisdiccional y en los más diversos campos de actuación estatal.

REFERENCIAS

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 Initiatives*. 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: Cambridge 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-egrid>>. 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.