

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

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

Carlos Alberto Molinaro

Professor No PPGD Da Faculdade De Direito Da Pontifícia
Universidade Católica Do Rio Grande Do Sul - PUCRS
Email: carlos.molinaro@puers.br

Augusto Antônio Fontanive Leal

Doutorando em Direito pela Pontifícia Universidade Católica do Rio Grande
do Sul (PUCRS). Mestre e graduado em Direito pela Universidade de Caxias do Sul (UCS).
Advogado.
Email: augustoafleal@gmail.com

RESUMO

O presente artigo tem por objetivo realizar uma análise explicativa sobre o conteúdo das novas tecnologias denominadas de Big Data e *machinelearning*, 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 *machinelearning*; b) uma exposição do aprimoramento advindo da utilização do Big Data e de técnicas de *machinelearning* 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 *machinelearning* 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.

*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.*

INTRODUÇÃO

Devido aos constantes impactos gerados sobre a biodiversidade em um cenário global, já é possível antever uma calamidade ecológica *in fieri*. A complexidade dos problemas ambientais está enredada em questões de difícil compreensão, sobretudo versando sobre os riscos ambientais. De igual forma, as questões relativas aos princípios da prevenção e da precaução também são acesso dificultoso diante das distintas dependências encontradas na biodiversidade, a partir da estrutura do meio ambiente e das diversas relações e consequências desencadeadas por cada ato humano. Não obstante, parece que com a utilização de recursos tecnológicos pode ser possível alcançar meios mais eficientes de combater a degradação ambiental não somente em níveis locais, mas de modo global. Neste sentido, a incontestável evolução exponencial da tecnologia permitiu que atualmente se pudesse contar com as boas práticas do Big Data e do *machinelearning*.

É por meio deste estudo que se buscará estabelecer uma relação entre inovações tecnológicas de grande porte, como é o caso do Big Data e dos algoritmos de *machinelearning*, viabilizando um ferramental de proteção ambiental com maior precisão que somente uma grande quantidade de dados e uma portentosa capacidade de processamento poderiam oferecer. Para tanto, o presente ensaio está dividido em dois capítulos. No primeiro, se tentará uma abordagem sobre a metódica do Big Data e as técnicas de *machinelearning*, propiciando-se sua compreensão. No segundo, buscar-se-á efetuar uma aproximação destas inovações tecnológicas na proposta de preservação ambiental, com o estudo de programas existentes que utilizam estas tecnologias para aumentar e melhorar a conservação da natureza.

1. BIG DATA E MACHINE LEARNING: ABORDAGEM E CONCEITOS

Em um mundo conectado, dados são constantemente gerados. Vários são os aparatos tecnológicos que possibilitam esta geração crescente de dados e que torna possível o estabelecimento de padrões, uma abrangente compreensão do mundo ou, até mesmo, a capacidade de previsão de acontecimentos. Com o crescimento exponencial tecnológico, a velocidade de processadores e a capacidade de armazenamento de larga escala tornaram possível a volumosa figura do Big Data e o processamento

destes dados possibilitando uma compreensão que, sem técnicas de *machinelearning*, não seria possível. Para dispor sobre estas temáticas, este capítulo será dividido em duas seções, sendo que na primeira se estudará o Big Data e as oportunidades que são proporcionadas pelo seu crescimento incessante; na segunda, será estudada a revolução cognitiva por meio de técnicas de *machinelearning* de modo a possibilitar serem encontrados padrões na enorme quantidade de dados gerados, permitindo a compreensãodessa ferramenta tecnológica em desenvolvimento.

1.1 Big Data: olhando para um novo mundo de oportunidades

Por certo que a utilização de dados tenha aumentado pelas sofisticações oferecidas pela tecnologia e seus incessantes avanços, porém não se trata de uma tarefa decorrente do surgimento de computadores. Há muito tempo são procurados padrões em dados colhidos, como é o caso da busca efetuada por um fazendeiro por padrões em uma cultura, ou de um político na opinião de um eleitor (WITTEN; FRANK; HALL, 2011, p. 4). Dados sempre foram armazenados de alguma forma na vivência humana, para as mais diversas questões, seja em um aspecto individual como na memória, seja em documentos e registros. Se em um determinado momento histórico o armazenamento de dados era realizado por um caçador que procurava compreender um padrão no comportamento migratório da sua caça, isso pode ter passado para o armazenamento de dados para facilitar o comércio.

Porém, em um ambiente de desenvolvimento tecnológico, pode-se dizer estar ocorrendo uma sobrecarga de dados. Como referem Witten, Frank e Hall, a quantidade de dados no mundo e na vida humana está crescendo e não há um final para isso (WITTEN; FRANK; HALL, 2011, p. 3). Segundo os autores, isso ocorre devido à onipresença de computadores na vida cotidiana, que por meio de não tão caros sítios de armazenamento, como é o caso de armazenamento online, trazem consigo a facilidade de salvar aquilo que anteriormente teria sido dispensado (WITTEN; FRANK; HALL, 2011, p. 4). Os dados simplesmente vão sendo armazenados ao longo dos dias, por meios virtuais, em um cenário de facilidade tal que a qualquer momento pode ser a memória de qualquer dispositivo ampliada para comportar um maior número. Todavia, a geração de dados não se atém a tão somente a vida comum de cada ser humano, mas está em inúmeras atividades corporativas, governamentais, econômicas e sociais.

Demonstrando a capacidade de desenvolvimento de processadores (CPU) ainda em 1965, Gordon Moore comprovou que o número de *transistors per squared* dobraria a cada ano e, no ano de 1975, revisando a sua teoria, Moore dobrou este prazo para dois anos (BELL, 2016). Conjugado a isso, há um declínio do custo para armazenamento de dados (KATZ, 2013, p. 916), proporcionando um campo fértil para o desenvolvimento do que se cunhou de Big Data.

Quão grande é o Big Data?

Katz remete a um alvo subjetivo e sempre em movimento, de modo que um dos melhores caminhos para compreender este conceito seria um olhar para trás (KATZ, 2013, p. 916). O aprimoramento contínuo de sistemas de memórias em conjunto com uma capacidade de processamento se expandido cada vez mais demonstra que em um *continuum*, há uma possibilidade de armazenamento simplesmente incapaz de ser compreendida em dispositivos eletrônicos comuns e isolados.

Contudo, *big data* permanece como um conceito abstrato, de modo que apesar do amplo reconhecimento da sua importância, pairam diferentes opiniões sobre sua definição, podendo ser conceituado, em geral, como *datasets* que não podem ser percebidos, adquiridos, administrados e processados por tradicionais softwares e hardwares dentro de um tempo tolerável (CHEN et al., 2014, p. 2). Segundo a definição trazida pelo International Data Corporation (IDC), Big Data descreve uma nova geração de tecnologias e arquiteturas, desenhada para economicamente extrair valores de imensos volumes de uma larga variedade de dados, por meio de uma altamente veloz captura, descoberta e/ou análise (GANTZ; REINSEL, 2011, p. 6). Por isso, é importante esclarecer que Big Data não trata somente do conteúdo original armazenado ou sendo consumido, mas também da informação envolta deste consumo, o que pode ser exemplificado pela forma como os *smartphones* produzem uma fonte de dados adicionais que estão sendo capturados, como é o caso da localização geográfica, mensagens de textos e histórico de navegadores de internet (GANTZ; REINSEL, 2011, p. 6).

Gerados pelos mais diversos meios tecnológicos, os dados em uma quantidade absurdamente difícil de ser percebida, apreendida e mensurada não atendem a um crescimento linear, mas sim exponencial. Como explicado por Katz, há um desenvolvimento exponencial em uma sinergia entre os conceitos de Lei de Moore, Big Data, e a revolução da inteligência artificial, de modo que cada duplicação da velocidade de processadores, da

divisão pela metade do preço para armazenamento de dados e os avanços do *machinelearning* abrem uma imensidão de possibilidades em uma perspectiva não linear (KATZ, 2013, p. 922).

Neste cenário horizontalmente transversal, Big Data se forma a partir de diversas fontes, podendo incluir variados tipos de dados em uma quantidade difícil de ser absorvida. Avulta nesta grande quantidade, as novas fontes de dados presente nas mídias sociais em que os usuários, por vezes até inconscientemente, acabam gerando um grande volume de dados que pode ser utilizado em variadas formas e que vão se expandindo em rápidas taxas algorítmicas. Como descrevem Gantz e 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. *(Os grandes dados são uma seção transversal horizontal do universo digital e podem incluir dados transacionais, dados armazenados, metadados e outros dados que residem em arquivos ridiculamente grandes. Mídia / entretenimento, cuidados de saúde e vigilância por vídeo são exemplos óbvios de novos segmentos de grande crescimento de dados. Soluções de redes sociais como Facebook, Foursquare e Twitter são as novas fontes de dados mais recentes. Essencialmente, eles criaram sistemas onde os consumidores (conscientemente ou inconscientemente) estão fornecendo fluxos de dados quase contínuos sobre si mesmos, e graças ao “efeito de rede” de sites bem-sucedidos, os dados gerados podem expandir em taxas logarítmicas rápidas - Trad. Livre.)*

Porém, embora as mídias sociais representem uma grande contribuição na geração de dados, não se trata de uma única fonte. Além das plataformas de mídias sociais que produzem uma grande quantidade de dados pelas redes de perfis individuais ou por outros conteúdos produzidos, que tratam de blogs, fóruns de discussão de usuários, até mesmo “likes”, contém potenciais informações a serem usadas, podendo ser mencionadas como fontes os dados produzidos por atividades como o conteúdo mecanicamente gerado na forma de arquivos de registro de dispositivos,

na função de internet das coisas (*internet of things*). Outras fontes de dados podem ainda ser encontradas em serviços de softwares, aplicações em nuvem e as até então um tanto quanto inexploradas, são os dados de registros médicos e correspondências com clientes (RUNCIMAN, 2014, p. 1). Não é à toa que, de acordo com a IBM (2017), diariamente são criados 2,5 quintilhões de bytes de dados. É tanto que 90% dos dados no mundo hoje foram criados nos últimos dois anos, tratando-se de dados que vieram de todo lugar: sensores de informação sobre o clima, postagens em mídias sociais, fotos digitais, vídeos, registros transacionais, registros de sinais de GPS de celulares, dentre inúmeras outras formas.

As possibilidades existentes a partir da utilização do Big Data são as mais diversas e representam um fértil campo para aprimoramento da vida humana e de todas as espécies de vida da terra. É preciso dispor de meios para a utilização destes dados que contemplem sua identificação, verificação, compreensão e utilização. Basu e Hall, demonstrando a capacidade da utilização deste recurso tecnológico, citam a presença do Big Data no *Square Kilometre Array*, um rádio telescópio com tamanho capaz de torná-lo cinquenta vezes mais sensível e dez mil vezes mais rápido que qualquer outro, que irá gerar por dia a quantidade de dados existentes na internet, tornando astrônomos capazes de perceber qualquer sinal na galáxia e olhar mais profundamente pelo espaço como nunca antes (BASU; HALL, 2014, p. 46); de acordo com esses mesmos autores, o Big Data é parte de uma revolução no mundo doméstico, social e dos negócios, pois assim como o maciço rádio telescópio *Square Kilometre Array* vai gerar Big Data para habilitar um melhor entendimento da história e lugar no espaço do homem, existem grandes oportunidades para incrementar o uso de Big Data para entender melhor o futuro ambiente humano e as possibilidades no planeta (p. 51).

Em decorrência de um desenvolvimento tecnológico exponencial, com processadores cada vez mais potentes e a diminuição do custo de armazenamento de dados, possibilitou-se uma geração de dados de gigantesca quantidade e que cresce diariamente, tendo como fonte a criação de dados os conscientes ou inconscientes usuários de mídias sociais ou demais dispositivos conectados, o registro realizado mecanicamente de diversos aparelhos com conexão (internet das coisas), o armazenamento em nuvem e diversos outros dados que podem vir a ser explorados. Esses dados em larga escala que ajudam a construir o significado de Big Data oferecem inúmeras possibilidades e oportunidades para revolucionar

positivamente a vida humana na Terra e, como é no caso da problemática ambiental, propiciando um campo fértil na proteção ambiental.

Porém, não basta somente que se reconheça a existência do Big Data. Sendo necessários meios para que estes dados possam vir a ser utilizados, isto é, é preciso descobrir ou canalizar a grande quantidade de dados por meios que tornem possível a sua utilização. Como demonstrado, trata-se de uma quantidade de dados absurdamente imensa e não há como o intelecto humano fazer este serviço sem um ferramental tecnológico capaz de minerar estes dados, selecionar o que se está buscando e verificar, por meio de alguma análise, aquilo que se quer. É neste sentido que surge o *machinelearning* uma revolução no campo da inteligência artificial, como se demonstrará no próximo subcapítulo.

1.2 Revolução cognitiva, *machinelearning* e a mineração de dados

O colossal volume de dados que dia após dia é gerado, para ser utilizado de acordo com uma proposta de alcançar um aprimoramento na vivência humana pelas oportunidades que existem e virão a existir com o Big Data depende de um aparato tecnológico que dê conta não somente de processar e interpretar herculeamente uma dada quantidade de dados, mas que realize tarefas de escolha e possa, efetivamente, aprender com a análise gerada para possibilitar respostas de análise de comportamento, identificando padrões que possam ser utilizados.

Embora um ferramental deste porte possa ser dirigido por algum ser humano, é impossível conceber que a mente humana possa trabalhar com tamanho volume de dados como demonstrado naquilo que se cunhou de Big Data. Como evidenciado por Witten, Frank e Hall, há uma grande lacuna (*gap*) entre geração de dados e o entendimento humano destes dados (WITTEN; FRANK; HALL, 2011, p. 4). Parece certo que, enquanto o volume de dados segue aumentando, a aptidão humana para compreender isso diminui de maneira alarmante, fazendo com que uma grande quantidade de dados permaneça não utilizada (WITTEN; FRANK; HALL, 2011, p. 4). Esses dados, com potencial utilidade, raramente vêm a ser identificados ou utilizados, permanecendo em um estado de invisibilidade. A mineração de dados (*data mining*) não é um recurso tão recente, uma vez que economistas ou engenheiros de comunicação, por exemplo, desde muito trabalham com a ideia de encontrar padrões em dados automaticamente, identificando, validando e utilizando para previsões. O que há de novo é

um incremento assombroso nas oportunidades para que sejam encontrados estes padrões em dados que advém do insaciável crescimento (WITTEN; FRANK; HALL, 2011, p. 4).

Por certo que, de acordo com Katz, a capacidade de armazenamento de dados e a velocidade de processadores não são suficientes, por si só, de gerar o conjunto de percepções agregadas que levam a uma maior produtividade e inovação. Segundo o autor, além destas duas inovações tecnológicas, uma revolução da inteligência artificial deve ser observada como um terceiro pilar em uma nova era de produtividade (KATZ, 2013, p. 918). É neste contexto que surge a utilização de técnicas de *machinelearning*, (que neste texto será utilizado o termo original que pode ser compreendido em português como aprendizagem computacional ou aprendizagem automática) aparecendo como uma ferramenta capaz de encontrar e descrever padrões em dados (WITTEN; FRANK; HALL, 2011, p. 5).

Contudo, há um conteúdo metafórico em dizer que um computador pode aprender, aparentando uma expressão meramente ilustrativa. Analisando cinco definições sobre aprender: a) ter conhecimento de algo pelo estudo, experiência, ou aprendendo; b) tornar-se ciente por uma observação ou informação; c) memorizar; d) ser informado ou averiguar algo; e) receber instrução, perguntam Witten, Frank e Hall sobre como se poderia saber se um computador aprendeu alguma coisa? Ou se um computador se tornou ciente de algo? A questão que diz respeito sobre a possibilidade de um computador estar ciente ou consciente é uma questão atual da filosofia. Porém, os últimos três aspectos (itens “c”, “d” e “e”), poderiam ser muito aproximados do significado de *machinelearning* (WITTEN; FRANK; HALL, 2011, p. 7). Caso em que, ao invés de buscar uma réplica do processo cognitivo humano em computadores que estivera baseada na crença de criar artificiais versões da funcionalidade cerebral humana, ficando adstrito a um suposto processo inteligente de produção, está a possibilidade alternativa de se obter resultados que processos automáticos alcançam, sendo precisos, apropriados, auxiliares e úteis, podendo ser considerados inteligentes, ainda que não advenham de uma réplica artificial de um processo cognitivo humano (SURDEN, 2014, p. 98-99).

Deste modo, a aprendizagem poderia ser atribuída às coisas quando estas mudam seu comportamento de um modo que possam ter um melhor desempenho no futuro, tratando-se muito mais de uma questão de desempenho do que de conhecimento (WITTEN; FRANK; HALL,

2011, p. 7). Como referido por Surden, operando bem, algoritmos de *machinelearning* podem produzir resultados automáticos aproximados do que realizariam uma pessoa igualmente situada (SURDEN, 2014, p. 90). Assim, a utilização de dados, pela mineração e posterior verificação com a finalidade de encontrar um padrão de comportamento que possa ser utilizado não somente para compreender algo ocorrido ou que esteja ocorrendo, mas para dispor sobre uma previsão, diante de uma grande quantidade de dados (Big Data), se daria por técnicas de *machinelearning*.

Com o exemplo da classificação automática de e-mails como spam, Surden busca demonstrar que algoritmos de *machinelearning* podem se desenvolver cada vez mais assim receber uma maior quantidade de dados, tendo por meta a construção de um modelo de um complexo fenômeno que irá capacitar um computador de realizar classificações precisas (SURDEN, 2014, p. 92). Com base nisso, pode-se dizer que computadores podem vir a ter resultados “inteligentes” em tarefas sem a intervenção da cognição humana (SURDEN, 2014, p. 95).

Em suma, pelo processo de aprendizagem se teria o aprimoramento de um comportamento pelo estudo de experiências (RUSSEL; NORVIG, 2010, p. 693). As técnicas de *machinelearning*, operariam aprendendo a partir da experiência e desenvolvendo uma melhora de seu desempenho pelo tempo, o que contribuiria sobremaneira na medida em que fosse implementada nos mais diversos setores, como econômicos, sociais, ambientais, capacitando uma compreensão e oportunizando previsões.

Há que se dizer, porém, em um aspecto problemático presente nas previsões. Se o objetivo da utilização de técnicas de *machinelearning* gestão pautados na análise de dados para obter resultados aproximados no futuro, em cenários ainda não vistos (ou nunca vistos), pode-se esbarrar no problema da generalização (SURDEN, 2014, p. 105). Surden reparte o problema da generalização em três aspectos: a) se eventos futuros dependerem de fatos únicos ou não incomuns, a previsão acabaria dificultada; b) deve-se evitar uma sobregeneralização (*overgeneralization*), pois é indesejável que algoritmos de *machinelearning* detectassem dados idiossincráticos ou preconceituosos que, obviamente, não teriam condições de fazer uma previsão; c) é o caso, por exemplo, de previsões legais, quando algum algoritmo de *machinelearning* encontre dificuldades de capturar sutis fatores de grande relevância (SURDEN, 2014, p. 105-106).

O problema da generalização poderia fazer cair por terra uma previsão que se buscou encontrar quando algum fator chave não esteja

presente nos dados anteriores, quando existir algum caso em que os dados estão dotados de falsas percepções como, por exemplo, pré-juízos que impediriam uma análise correspondente com a realidade e, ainda, quando existissem fatores sutis que dificilmente seriam captados pela técnica de *machinelearning* utilizada. Para atuar sobre estes fatores, é importante que os dados utilizados sejam filtrados em um critério tal que se possa mensurar determinadas características. Também, determinados problemas existentes na possível ocorrência de atipicidades ou de fatores de difícil percepção podem ser aprimorados ao longo da utilização do sistema de *machinelearning* desenvolvido.

Por isso, para maximizar a potência de técnicas de *machinelearning*, em uma teoria de aprendizagem computacional que esteja pautada na interseção entre inteligência artificial, estatística e teoria da ciência computacional, demonstra-se um princípio subjacente de que qualquer hipótese que seja seriamente errada quase certamente seria encontrada, após um pequeno número de exemplos, pois estaria fazendo uma errada previsão, de modo que qualquer hipótese consistente com um largo sistema de exemplos demonstraria um improvável erro, estando provavelmente aproximadamente correta (*probablyapproximatelycorrect*) (RUSSEL; NORVIG, 2010, p. 714). Para tanto, a capacidade de processar uma grande quantidade de dados, testando hipóteses com uma aproximação cada vez maior de uma resposta correta e coerente, teria a função de eliminar problemas decorrentes de uma generalização, até mesmo quando para uma previsão se dependesse de um fato ainda não ocorrido, pois uma série de experiências levaria, após um dado período, na descaracterização deste problema pelo conteúdo de circunstâncias até então analisadas. Concluindo, Flach concebe *machinelearning* como sendo o estudo sistemático de algoritmos e sistemas com a finalidade de melhorar seu conhecimento ou desempenho pela experiência (FLACH, 2012, p. 3). Ao longo do tempo, técnicas de *machinelearning* tornam-se mais proficuas e permitem, diante de uma grande quantidade de dados, um aprimoramento de diversos setores da sociedade, revolucionando a vida humana.

A utilização de *machinelearning* que, por meio de imagens geradas diariamente por satélites radares que monitoram águas costeiras para detectar manchas de óleo, permite treinar um sistema detector de contaminação, com exemplos sobre derramamentos para cada utilizador mediante seus próprios propósitos é apenas um exemplo sobre como esta tecnologia pode auxiliar na preservação ambiental. No caso, ao encontrar

possíveis manchas de óleo, primeiramente o sistema processa e normaliza a imagem, identificando a mancha suspeita. Após, são determinados os atributos de cada mancha de óleo, como a região, o tamanho e a intensidade e, ao final, aprendizagens de padrões são utilizadas para os atributos resultantes (WITTEN; FRANK; HALL, 2011, p. 23).

Como afirmado por Flach, existem grandes chances de que algoritmos de *machine learning* já estejam sendo usados por um alto número de pessoas, ainda que estas não saibam disso (FLACH, 2012, p. 1). De igual forma, Russel e Norvig referem a existência de uma atuação por trás das cenas da inteligência artificial, como no caso da automática aprovação de compra de cartão de crédito. Estas mudanças que são incorporadas acabam modificando a vida diária das pessoas, ainda que muitas não se deem conta disso. Apesar de estar em desenvolvimento, pode-se esperar que um médio sucesso da inteligência artificial viria a afetar todas as pessoas em suas vidas (RUSSEL; NORVIG, 2010, p. 1051).

O certo é que o mundo como é conhecido vem mudando diariamente com o exponencial avanço tecnológico de processadores, técnicas de armazenamento de memória e incremento de algoritmos de *machine learning*. Oferece-se, assim, um vasto campo de possibilidades e oportunidades para melhorar a vida humana em diversos aspectos, dentre os quais e seguramente um dos mais importantes, está a preservação da natureza como um todo. Inclusive, algumas técnicas já vêm sendo desenvolvidas e servem como exemplos sobre como a tecnologia do Big Data e de *machine learning* podem trazer efetividade para assegurar um meio ambiente ecologicamente equilibrado, tanto em escala local como global. É o que se demonstrará no próximo capítulo.

2. A FUNÇÃO DA TECNOLOGIA DO BIG DATA E AS TÉCNICAS DE MACHINE LEARNING NA PRESERVAÇÃO AMBIENTAL

A utilização de algoritmos de *machine learning* aliados a uma capacidade portentosa de armazenamento de dados representam conjuntamente em significativas mudanças nos mais variados aspectos da vida humana. Por isso, tanto Katz (2013) quanto Surden (2014) buscam relacionar a utilização de *machine learning* para o melhor desempenho de tarefas legais, o primeiro autor, trabalha em uma nova era de previsão legal quantitativa que viria a auxiliar advogados, estudantes e faculdades de Direito (p. 912-913); já o segundo, adverte que não é insignificante,

número de tarefas legais que podem ser parcialmente automatizadas usando técnicas de inteligência artificial (p. 88).

Na busca pela preservação da natureza isso não é diferente (!).

O desenvolvimento tecnológico amplia o ferramental que serve tanto para encontrar os problemas ambientais como para alcançar soluções. Tal perspectiva pode vir a trazer uma maior efetividade ao direito fundamental ao meio ambiente e suas determinações previstas no artigo 225 da Constituição Federal brasileira (BRASIL, 1988), como é o caso da defesa e preservação de um meio ambiente ecologicamente equilibrado, tendo em vista a realidade de um assunto *res maximimomentique* que levou a estarem constitucionalizados os fundamentos de proteção ambiental e do incremento de sua qualidade (MILARÉ, 2014, p. 160-161). Todavia, esta preocupação com a realidade natural não está adstrita a uma mera positivação. Para além disso, o meio ambiente atinge os níveis de um direito fundamental que conduz à formulação de um princípio de primariedade do ambiente, reconhecendo-se a ilicitude de se tratar o meio ambiente como subsidiário, acessório, menor ou desprezível (BENJAMIN, 2010, p. 118).

Atente-se que a proposta de um projeto como o Global Forest Watch (<http://www.globalforestwatch.org/>), que consiste em um sistema de alerta e monitoramento de florestas online que oferece ferramentas para que qualquer pessoa do mundo possa ser capaz de gerir e conservar as floretas, é um claro exemplo da utilização de ferramentas tecnológicas para a preservação ambiental. O trunfo tecnológico de um projeto desta magnitude não está somente pautado em uma gigantesca coleta de dados de nível global (Big Data), mas também em uma união de pessoas do mundo todo no fornecimento de dados e utilização de sistemas de alerta que viabilizam uma maior proteção florestal.

O supracitado artigo 225 da Constituição Federal impõe, tanto ao Estado como à coletividade, a preservação e defesa do meio ambiente ecologicamente equilibrado. Justamente por isso, conforme Sarlet e Fensterseifer, subsiste uma diferença entre este direito e os ditos direitos fundamentais de primeira dimensão, pautados nos direitos liberais, bem como do princípio da igualdade, assentado em direitos sociais de segunda dimensão, concluindo pelo pilar normativo-axiológico do direito fundamental ao meio ambiente como princípio de deverde solidariedade (SARLET; FENSTERSEIFER, 2013, p. 47). Com isso, o próprio sistema cooperativista em projetos de alta ponta tecnológica, utilizando Big Data, como o Global FlorestWatch possibilitam meios para que os cidadãos e o

próprio governo brasileiro tenham condições de cumprir com seu dever de proteção ambiental presente no conteúdo normativo do artigo 225 da Constituição Federal. Além das regulamentações internas, deve-se levar em conta também o cumprimento de tratados internacionais que dizem respeito às mais diversas questões de nível global, demonstrando que não só para uma execução eficaz de normas internacionais, a utilização de Big Data e *machinelearning* servem para estabelecer tratados justos e que tenham um conteúdo preciso e necessário de regulamentação.

Mais profundamente, porém, se buscará demonstrar nos seguintes subcapítulos como significativos impactos tecnológicos como Big Data e técnicas de *machinelearning* podem influenciar e redirecionar os esforços na preservação ambiental por meio de um aprimoramento geral das ferramentas a serem utilizadas.

2.1 BirdCaste a migração de pássaros em tempo real

BirdCast (<http://birdcast.info/>) é um programa desenvolvido com a finalidade de, pela primeira vez, viabilizar previsões de migrações de pássaros, estabelecendo condições para saber quando eles migrarão, para onde eles migrarão e quanto eles voarão. O projeto, desenvolvido pelo The Cornell Lab of Ornithology, propõe que o conhecimento do comportamento migratório auxiliará em critérios de conservação, como é o caso do posicionamento de turbinas eólicas ou até na identificação de que em determinada noite algum alto prédio poderá ter que deixar suas luzes acesas prevenindo a morte de milhões de pássaros. Porém, a importância do BirdCast vai além disso. Higuchi, ao mencionar ter colaborado com cientistas americanos e asiáticos no rastreamento por satélite da migração de pássaros por vinte anos, demonstra a importância do conhecimento migratório de pássaros para a conservação do meio ambiente global (HIGUCHI, 2012, p. 3). Refere o autor que pássaros migratórios encontram vários problemas durante a migração, o que inclui a destruição de habitats, poluição química, caça e colisões com aviões (HIGUCHI, 2012, p. 3).

Há uma ligação entre pássaros migratórios e questões futuras sobre a relação o meio ambiente e atividades humanas. A cada outono e primavera, um elevado número de pássaros visitam diversas áreas em suas rotas migratórias, como sem comida não há migração, obtém comida e outras questões significativas nestas áreas. Conseqüentemente, pássaros migratórios possuem um importante papel na manutenção de ecossistemas

em diferentes áreas em decorrência de seus hábitos alimentares, de modo que se somente poucos pássaros migratórios existirem, a população de insetos poderia aumentar drasticamente em florestas causando um desequilíbrio na natureza, assim como algas poderiam crescer densamente em lagos, poluindo a água (HIGUCHI, 2012, p. 11). Por vezes, estes impactos negativos quando de uma mudança na migração de pássaros não são bem compreendidos. A diminuição da população de pássaros migratórios em determinado habitat pode acarretar na deterioração de um ecossistema em outra área. Segundo Higuchi, a destruição de florestas tropicais no sudeste asiático diminui o número de pássaros migratórios tropicais que ali passam o inverno e migram para o Japão na primavera, gerando um desequilíbrio na natureza de ecossistemas florestais no Japão. Também, a destruição de planícies de marés no Japão pode diminuir o número de muitas aves limícolas, provocando uma deterioração de ecossistemas de zonas húmidas nas Filipinas, Austrália e Rússia (HIGUCHI, 2012, p. 12). Por isso, questões como estas representam um desafio para o desenvolvimento da ornitologia e da ecologia, evidenciando que a conservação de pássaros não se volta somente para sua preservação, mas também para a manutenção de ecossistemas, levando à conservação do meio ambiente global (HIGUCHI, 2012, p. 12).

O programa BirdCast disponibiliza modelos precisos de migração de pássaros, permitindo que pesquisadores compreendam aspectos comportamentais da migração, como o tempo de migração e as vias de migração respondem as mudanças climáticas, e quais ligações existem entre o tempo da migração e as subsequentes mudanças no tamanho populacional. Dentre as fontes de dados do programa estão as chamadas de vôo que permitem identificar a composição de espécies de migração noturna quando de sua ocorrência fornecendo informações em tempo quase real, os radares de vigilância meteorológica, que podem ser utilizados para observar e rastrear os movimentos de pássaros e, por fim, o revolucionário eBird.

Iniciado em 2002, pelo The Cornell Lab of Ornithology e o National Audubon Society, eBird é um programa que oferece uma lista em tempo real para comunicar relatórios e acessar informações sobre pássaros, fornecendo dados fontes para informações básicas na abundância de pássaros e a distribuição em escalas espaciais e temporais. O programa eBird é um dos maiores e mais rápidos acumuladores de recursos de dados de biodiversidade que existem, tendo em maio de 2015, mais

de nove milhões de relatórios de observações de pássaros através do mundo. Valendo-se destas fontes de dados, inclusive com o portentoso recurso do eBird, o programa BirdCast propõe o desenvolvimento de duas técnicas de *machinelearning*, *CollectiveGraphicalModels* (CGMs) e *Semi-ParametricLatentProcessModels* (SLPMs). As duas técnicas têm por objetivo, conjuntamente, reconstruir e prever o comportamento de quatrocentas espécies de pássaros migratórios na América do Norte, resultando em um modelo capaz de identificar as complexas condições que governam a dinâmica do comportamento migratório. Além disso, BirdCast busca o desenvolvimento de dados interoperáveis para sintetizar observações de pássaros, chamadas de voo, dados de radares e dados de múltiplas fontes, como imagens de satélite, meteorologia e dados de populações humanas. Contando, ao final, com um sistema de dados em web para comunicar previsões migratórias geradas pelo BirdCast para o público.

Com a utilização de técnicas de *machinelearning* em conjunto com uma imensa fonte de dados cumulativa, BirdCast possibilita a compreensão do comportamento migratório de pássaros, fornecendo valiosas informações para aumentar significativamente a preservação ambiental em escala global.

2.2 Cadeia logística global e a identificação de pontos centrais de espécies ameaçadas

A cadeia logística representa a trajetória de um produto, desde concepção como um suprimento, passando pelo estágio de produção em um processo de industrialização e alcançando, ao final, a distribuição, que atende à demanda dos consumidores. Não raro, este processo se dá em um nível globalizado, gerando uma interligação de diversas partes neste complexo processo de produção. Companhias que agem sobre áreas críticas para espécies ameaçadas, no entanto, tendem a continuar com a sobre-exploração agravando ainda mais a problemática retratada.

A perda da biodiversidade e, mais precisamente, ameaças locais às espécies, estão sendo conduzidas pela atividade econômica e pela demanda de consumo (LENZEN et al., 2012, p. 111). Justamente por isso, 30% de espécies ao redor do mundo estão ameaçadas em decorrência do comércio internacional (LENZEN et al., 2012, p. 109). Compreender isso demonstra que não se tem como agir somente em nível local, em torno de

uma regulação que não atue somente sobre países em desenvolvimento que realizam exportações em elevado número, existindo a necessidade de que regulações atinjam países onde o consumo atua ativamente na criação de uma gravosa pegada ambiental colocando em ameaça inúmeras espécies.

A questão está justamente no cenário que envolve a identificação de pontos centrais de espécies ameaçadas a partir de uma cadeia logística global. E é justamente este o trunfo de uma pesquisa realizada pautada na construção de um mapa que combina diversos mapas de ocorrência de uma série de espécies ameaçadas. Com estes dados espaciais de uma pegada ambiental, é possível saber que países e quais categorias de consumo ameaçam habitats em vários pontos críticos (MORAN; KANEMOTO, 2017, p. 1). A pesquisa trouxe resultados notáveis, como é o caso da pegada ambiental sobre espécies terrestres proveniente dos Estados Unidos não somente sobre o sudeste asiático ou Madagascar, como era esperado, mas também em pontos críticos ao sul da Europa, no Sahel, na costa leste e oeste do sul do México, pela América Central e Ásia Central e no sul do Canadá. Apesar de haver uma atenção redobrada sobre a floresta Amazônica, o estudo revelou que a pegada dos Estados Unidos no Brasil é maior no planalto, ao sul, do que na Amazônia. A profunda pegada dos Estados Unidos sobre a biodiversidade ao sul da Espanha e de Portugal, em relação ao impacto no número de espécies de pássaros e peixes ameaçados, é digna de nota, em decorrência de estas áreas serem raramente considerada como um ponto crítico de espécies ameaçadas (MORAN; KANEMOTO, 2017, p. 1). Pela pesquisa também foi possível observar pontos críticos de ameaças para vários países consumidores majoritários e aproximar em cada particular região afetada pelo consumo. É o caso do impacto do consumo dos Estados Unidos na África do Sul, do consumo da Europa na África, e do sudeste da Ásia pelo consumo japonês (MORAN; KANEMOTO, 2017, p. 1).

Pela utilização do método de pontos críticos, tornou-se possível identificar que áreas em que a ameaça à biodiversidade é predominantemente conduzida por um pequeno número de países. Com isso, identificando as regiões em que determinados países estejam pressionando a biodiversidade, capacita-se o início de uma direta colaboração entre produtores e consumidores, atendendo às regulamentações internacionais, a fim de mitigar os impactos ambientais nos locais identificados (MORAN; KANEMOTO, 2017, p. 3). Contabilizando espacialmente os impactos explícitos na biodiversidade de determinadas áreas, é possível ajudar a melhorar a produção sustentável, o comércio internacional e o consumo.

A responsabilidade pelos impactos ambientais deve ser repartida ao longo de toda a cadeia de logística, das indústrias aos consumidores finais. Sendo que os mapas fornecidos pela pesquisa auxiliam desde os produtores aos consumidores, incluindo os conservacionistas, na busca por soluções no rastreamento de pontos críticos para a biodiversidade (MORAN; KANEMOTO, 2017, p. 3).

A pesquisa realizada somente se tornou possível por meio de um massivo conteúdo de dados de diversas fontes. Os resultados foram calculados a partir dos três seguintes dados disponíveis (MORAN; KANEMOTO, 2017, p. 5):

1. IUCN RedListofThreatenedSpecies: contém a diversidade genética e os blocos de construção dos ecossistemas, informações sobre os status de conservação e distribuição e capacitam decisões informadas sobre a conservação da biodiversidade em nível local e global. Fornece informações taxonômicas, status de conservação e distribuição a respeito de plantas, fungos e animais avaliados globalmente.

2. BirdLifeInternational: trata-se de uma parceria de nível global, formada por diversos parceiros, sendo cada um uma organização ambiental não governamental, sem fins lucrativos. BirdLifeInternational busca por meio de uma rigorosa ciência, implementar programas de conservação para pássaros e toda a natureza. A parceria global permite um trabalho para a natureza em cada local em que há um membro estabelecido, estando este conectado nacionalmente e internacionalmente.

3. Eora MRIO: possui um banco de dados de *inputs-outputs* com o auxílio de satélites sociais e ambientais em 187 países em 15.909 setores.

A importância desta pesquisa que tem se valido de gigantescos bancos de dados demonstra como é preciso uma aproximação direta da legislação de nível internacional e local para providenciar uma mais precisa preservação do meio ambiente em sua completude. Identificar a cadeia logística e as ameaças à biodiversidade que ocorrem em determinada local podem levar a conceber tratativas entre os diversos países envolvidos, buscando uma responsabilização compartilhada e uma maior conservação da biodiversidade.

2.3 Programas governamentais tecnológicos: um exemplo a partir de ferramentas tecnológicas utilizadas pela Agência de Proteção Ambiental estadunidense

De acordo com Breggin e Amsalem, Big Data vem sendo utilizado por agências governamentais, organizações não-governamentais, e empresas privadas para proteger o meio ambiente, sendo que as autoras exemplificam por meio da melhoria na eficiência energética, da promoção da justiça ambiental, do rastreamento da mudança climática e do monitoramento da qualidade da água (BREGGIN; AMSALEM, 2014, p. 3). Não é à toa que, segundo constatam as autoras, no governo federal estadunidense, a maioria das entidades governamentais responsáveis pela proteção ambiental tem usado Big Data em seu trabalho. A utilização pode ser ilustrada pela análise de risco, pesquisa, execução, educação pública e capacitação.

Na U.S. Environmental Protection Agency (Agência de proteção ambiental dos Estados Unidos), pode-se citar o monitoramento de qualidade de ar que vem sendo elaborado no escritório de pesquisa e desenvolvimento, tendo como um dos exemplos o enfoque em novas tecnologias que auxiliem o governo a avaliar a camada de ozônio e a conformidade com os padrões de qualidade de ar nacional em conjunto com um método de fusão de dados que combine o resultado dos monitoramentos para criar previsões (BREGGIN; AMSALEM, 2014, p. 6).

Um outro exemplo é o desenvolvimento do CompTox, um programa de pesquisa em toxicologia computacional, que tem por propósito lidar com a ausência de dados sobre saúde e meio ambiente em milhares de produtos químicos (BREGGIN; AMSALEM, 2014, p. 7). O programa utiliza uma avançada pesquisa de integração entre biologia, biotecnologia, produtos químicos e ciência da computação para identificar importantes processos biológicos que podem ser desencadeados de produtos químicos, de modo que milhares de produtos químicos podem ter seus potenciais riscos avaliados em um curto espaço de tempo.

Outro exemplo é o The Emissions & Generation Resource Integrated Database (eGRID), um inventário de dados federais sobre usinas de energia e empresas de energia coletados, abrangendo as características ambientais de quase toda eletricidade gerada nos Estados Unidos. Dentre as características estão a emissão atmosférica de poluentes, a taxa da emissão e a quantidade de energia elétrica gerada em uma usina elétrica que

é transmitida e distribuída ao consumidor. Estas enormes quantidades de dados são agregadas por instalações, empresas e pelo Estado (BREGGIN; AMSALEM, 2014, p. 7).

Com o objetivo de unificar sistemas, combinando e conectando satélites e sistemas de observação localizados no solo ao redor do mundo, o The Global Earth Observation System of Systems (GEOSS) é um projeto em desenvolvimento por The Group on Earth Observations (GEO), uma parceria internacional entre governos e organizações. Intenta-se auxiliar a comunidade internacional a coordenar respostas emergenciais aos desastres naturais ou causados pelo homem, tais como incêndios florestais, monitorar mudanças climáticas e na gestão de recursos naturais (BREGGIN; AMSALEM, 2014, p. 8). Pelo GEOSS, há um aumento do entendimento dos processos da Terra e uma melhora das capacidades preditivas que sustentam a tomada de decisões, provendo acesso a dados, informações e conhecimento para uma ampla variedade de usuários.

A aproximação tecnológica encetada pela Agência de Proteção Ambiental demonstra a imprescindibilidade da utilização de recursos tecnológicos de alta ponta para uma eficaz conservação do meio ambiente. A complexidade dos diversos fatores que dizem respeito à biodiversidade pode ser compreendida por recursos computacionais que tenham a capacidade de armazenar uma grande quantidade de dados, bem como de processar os dados colhidos proporcionando entendimento para decisões a serem tomadas no âmbito governamental. Ainda, a regulamentação ambiental pode ter maior precisão pela análise de uma grande quantidade de dados, Big Data, em conjunto com o processamento destes dados que podem realizar previsões sobre inúmeras questões, tornando possível uma antecipação com a finalidade de prevenir a ocorrência de desastres ambientais, dentre outros fatores que afetem a biodiversidade.

CONCLUSÃO

A revolução proporcionada pela capacidade de processadores de computador conjuntamente com o considerável aumento de recursos de armazenamento de dados levou à atual era do Big Data. A velocidade na criação de dados em uma estrondosa quantidade que torna impossível sua análise por sistemas de gestão tradicionais, aumentou as possibilidades para uma maior compreensão de diversos fatores do mundo, como é o caso da economia, sociedade e da natureza. Em conjunto com o Big Data, o

advento de técnicas de *machinelearning*, possibilitando um processamento desta vultosa quantidade de dados capacita esta compreensão, fornecendo um ferramental tecnológico como nunca antes visto e trazendo nos desafios do desenvolvimento e aprimoramento desta tecnologia, possibilidades inúmeras para os mais variados setores da sociedade.

No presente ensaio, buscou-se aproximar a era do Big Data como forma de proteger e preservar o meio ambiente. A interligação da biodiversidade em diversas regiões demonstra o caráter difuso do bem ambiental e a compreensão de suas teias de conexões por meio de uma grande quantidade de dados ultrapassa as barreiras locais e permite uma aproximação dos fatos como eles ocorrem, no plano global, estatal, regional e local. Com isso, possibilita-se um justo diálogo entre as partes envolvidas para estabelecer uma equivalente responsabilização pela preservação ambiental, bem como traz matérias a serem reguladas no plano internacional e interno de cada país.

Os projetos existentes, sejam os não-governamentais formados por membros de vários lugares do mundo em uma cooperação pela preservação ambiental, sejam os governamentais, como é exemplificado pelos desenvolvimentos da Agência de Proteção Ambiental americana, demonstram como o Big Data vem auxiliando o governo na preservação da natureza, alcançando o posto de peça chave para decisões mais precisas. Concluindo-se que a utilização de Big Data e de técnicas de *machinelearning*, além de ser possível, já é um sistema utilizado para preservar o meio ambiente, pode-se afirmar que no plano interno brasileiro, a utilização destes recursos tecnológicos deve ser implementada com a finalidade de trazer eficácia ao direito fundamental ao meio ambiente ecologicamente equilibrado, nas tomadas de decisões políticas, nas políticas públicas de justiça ambiental, no exercício jurisdicional e nos mais diversos campos de atuação estatal.

REFERÊNCIAS

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.

EORAMRIO.*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. *PáginaWeb*. 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. *PáginaWeb*. Disponível em: <<http://www.birds.cornell.edu/page.aspx?pid=1609>>. Acesso em 30 abr. 2017.

THE IUCN RED LIST OF THREATENED SPECIES. *PáginaWeb*. 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). *PáginaWeb*. 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). *PáginaWeb*. 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.