



Esta obra está licenciada com uma Licença Creative Commons Atribuição 4.0 Internacional

ISSN2175-9596



## BIG DATA PARA O BEM-ESTAR SOCIAL: POSSIBILIDADES E RISCOS NA UTILIZAÇÃO DO BIG DATA EM POLÍTICAS PÚBLICAS

*Big Data para el Bienestar social: posibilidades y riesgos en la utilización del Big Data en políticas públicas*

*Big Data for social welfare: possibilities and risks in the use of Big Data in public policies*

**Carolina Henrique<sup>a</sup>**

<sup>(a)</sup> Observatório de Políticas Públicas, Direito e Proteção Social (OPPDPS).

### Resumo

Quando se trata de políticas públicas, há que se considerar sempre o objetivo direto que se pretende atingir, além do indireto, que é o bem-estar da sociedade como um todo. Atualmente, vive-se uma crise econômica, financeira e política que assola não apenas o Brasil. Além disso, as cidades ao redor do mundo vivem um movimento de crescimento populacional sem precedentes. De acordo com relatório da ONU de 2010, 70% dos habitantes do planeta viverão em metrópoles com mais de 100 mil habitantes em 2050. Tudo isso significa um aumento de todas as dificuldades de gerir questões sensíveis das cidades como educação, saúde e violência. Diante desse quadro, meios “tradicionais” de políticas públicas não se apresentam mais como eficazes, devendo o setor público, a exemplo do setor privado, acordar para as possibilidades apresentadas pelo Big Data. No entanto, não existe uma solução perfeita para qualquer problema, de forma que é preciso estar atento para os prejuízos e violações a direitos que podem ser gerados pela análise do Big Data, que podem por em risco os objetivos pretendidos pela política pública.

**Palavras-chave:** Big Data; Distorções; Políticas públicas.

### Resumen

*Cuando se trata de políticas públicas, hay que considerar siempre el objetivo directo que se pretende alcanzar, además de lo indirecto, que es el bienestar de la sociedad como un todo. Actualmente, se vive una crisis económica, financiera y política que asola no sólo a Brasil. Además, las ciudades alrededor del mundo viven un movimiento de crecimiento poblacional sin precedentes. Según el informe de la ONU de 2010, el 70% de los habitantes del planeta vivirá en*

*metrópolis con más de 100 mil habitantes en 2050. Todo esto significa un aumento de todas las dificultades de gestionar cuestiones sensibles de las ciudades como educación, salud y violencia. En este marco, medios "tradicionales" de políticas públicas no se presentan más como eficaces, debiendo el sector público, a ejemplo del sector privado, acordar las posibilidades presentadas por el Big Data. Sin embargo, no existe una solución perfecta para cualquier problema, por lo que hay que estar atento a los perjuicios y violaciones a derechos que pueden ser generados por el análisis del Big Data, que pueden poner en riesgo los objetivos perseguidos por la política pública.*

**Palabras clave:** Big Data; Distorsiones; Políticas públicas.

### **Abstract**

*When one thinks about public policies, one must always consider the imminent objective and the general objective that is the social good of the society. Nowadays we face a global economic, financial and political crisis. Besides that, cities around the world are living a movement of population grow without precedents in history. A UN report of 2010 stated that by 2050 70% of the planet inhabitants will live in cities with more than 100 thousands inhabitants each. All this means that the already existing problems in the cities such as education, health and violence will get amplified. To address this situation, "traditional" means of making public policy are insufficient. The public sector must, like the private sector, embrace the possibilities presented by the Big Data. However, there is no silver bullet so we must be alert for the risks and possible violations of fundamental rights that can be caused by Big Data analytics. Treated without care, they can put at risk the very objectives that the public policies intent to reach.*

**Keywords:** Big Data; Biases; Public policies.

### **INTRODUÇÃO**

Todo mundo que já jogou SimCity sabe como é difícil gerir uma cidade. Trata-se de um sistema complexo com muitas variáveis a serem consideradas e na qual os resultados muitas vezes são de difícil previsão.

De acordo com relatório da ONU de 2010, 70% dos habitantes do planeta viverão em metrópoles com mais de 100 mil habitantes em 2050. Isso significa um aumento de todas as dificuldades de gerir questões sensíveis das cidades como educação, saúde e violência.

Diferente do jogo, no entanto, na vida real não é possível afastar as dificuldades destruindo a cidade com terremotos e furações e começando do zero novamente. As questões são muito mais complexas de forma que as políticas públicas tradicionais não oferecem respostas satisfatórias.

O aumento da concentração populacional também ocasiona um aumento no volume de dados gerados nesses espaços urbanos de forma que a resposta passa a ser, cada vez mais, a utilização desses dados em políticas públicas na busca por cidades melhores.

Na medida em que governos e instituições sem fins lucrativos vêm descobrindo o Big Data para o bem-estar social ou para o desenvolvimento, como vem sendo chamado, vem surgindo novas possibilidades de uso dessa enorme quantidade de dados para os mais diversos fins.

No entanto, sem tirar o mérito dessas potencialidades, esse trabalho pretende demonstrar que a utilização do Big Data, e dos algoritmos que processam esses dados, não pode ser feita sem os devidos cuidados. Utilizado de forma descuidada, a mineração de dados pode ferir direitos como a privacidade e a igualdade e reproduzir desigualdades existentes. Como bem definiram Barrocas e Selbst (2016, p. 671):

“Approached without care, data mining can reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers, or simply reflect the widespread biases that persist in society. It can even have the perverse result of exacerbating existing inequalities by suggesting that historically disadvantaged groups actually deserve less favorable treatment”<sup>1</sup>.

Para tanto, em um primeiro momento, será explicado do que se trata e analisar as possibilidades de uso do Big Data para o bem-estar social. Após, em um segundo momento, será analisado o caso do sistema utilizado pela polícia americana, denominado Predpol, no qual se constatou resultados discriminatórios em face de grupos considerados como vulneráveis.

Por fim, este trabalho se volta para o panorama atual brasileiro, examinando as políticas públicas realizadas recentemente na cidade de São Paulo e no Rio de Janeiro e os riscos que elas podem trazer para direitos fundamentais, em especial para a privacidade e igualdade, assim como a questão da vigilância excessiva.

## **POSSIBILIDADES DE UTILIZAÇÃO DO BIG DATA**

O mundo gera diariamente 2,5 quintilhões de bytes sendo que atualmente, a cada um ano e meio, se gera a mesma quantidade de informação já criada pela humanidade desde o seu surgimento. Esse

---

<sup>1</sup> Tradução livre da autora: “Abordada sem cuidados, a mineração de dados pode reproduzir padrões de discriminação existentes, herdar prejuízos de antigos tomadores de decisão, ou simplesmente refletir as distorções que persistem na sociedade. Pode até gerar o resultado perverso de exacerbar desigualdades existentes ao sugerir que determinados grupos que sofrem desvantagens históricas na verdade merecem um tratamento menos favorável”.

volume incomensurável de dados é denominado de Big Data.

Tendo surgido em meados da década de 90 e cunhado pela NASA, descreve uma quantidade massiva de dados disponíveis, além das capacidades computacionais e analíticas por processos tradicionais. Através da análise e mineração desses dados se torna possível agregar informações de origens diversas de forma a relacioná-las e gerar conclusões, auxiliando de forma cada vez mais eficiente na tomada de decisões.

Entre as diversas fontes de coleta de dados pode-se apontar:

- (1) *Data Exhaust* – dados coletados de forma passiva de transações feitas por pessoas ao utilizarem serviços digitais como celulares, compras *online*, buscas na *internet*, entre outros;
- (2) Informação *online* – conteúdo da *internet* como *sites* de notícia ou interações nas redes sociais;
- (3) Sensores físicos – imagens de satélites ou câmeras de vigilância de paisagens, tráfego, desenvolvimento urbano entre outros;
- (4) Informações de cidadãos ou *Crowd-sourced*<sup>2</sup> – informação produzida ativamente ou submetida pelos cidadãos através de celulares (principalmente) como questionários, *hotlines*, etc.

Em meados dos anos 2000 o Big Data ganhou a mídia através da definição cunhada por Doug Laney, membro da Gartner, de seus três Vs: Volume, dados massivos de inúmeras fontes; Velocidade, capacidade e tempo de processamento sustentável ao objetivo e a necessidade de processamento em tempo real para diversos segmentos e; Variedade, diversas fontes como vídeos, fotos, *hashs*, transações financeiras etc.

A empresa SAS, que possui uma família de software SAS (*Statistical Analysis System*), utilizam mais duas dimensões para descrever o Big Data, além dos três Vs de Laney, que são a Variabilidade, inconsistência de fluxos de dados variados com picos sazonais, e a Complexidade no qual dispensa definição. Majoritariamente, porém, passou-se a ser adotado como definição os cinco Vs: Volume, Variedade, Velocidade, Veracidade e Valor.

---

<sup>2</sup> A palavra “*crowdsourcing*” se refere ao uso de atores não oficiais (“*the crowd*”) como uma fonte de informação, conhecimento e serviços, em oposição à prática comercial de “*outsourcing*”.

No entanto, apenas recentemente o Big Data efetivamente caiu nas graças da sociedade e das empresas, passando a ser denominado como “o novo petróleo” (Kohn, 2013; The Economist, 2017). O setor privado não demorou a descobrir diversas utilidades para a mineração massiva de dados, em especial na capacidade de gerar lucro diante de um tratamento mais individualizado ao seu consumidor. Conforme demonstrado pela pesquisa realizada pela consultoria em negócios norte-americana Bain & Company, as empresas que utilizam Big Data possuem cinco vezes mais chances de tomarem decisões mais rápidas do que seus concorrentes e duas vezes mais chances de obterem performance superior.

## **A UTILIZAÇÃO DO BIG DATA EM POLÍTICAS PÚBLICAS**

Considerando que dados estão sendo gerados pelas pessoas em tempo real, torna-se possível, como já percebido pelo setor privado, a sua análise em tempo real por computadores de alta performance, criando, como dito anteriormente, um enorme potencial de auxílio na toma de decisões. Se as possibilidades para as empresas já são inúmeras, esse número aumenta quando se trata do setor público.

Quando se trata de políticas públicas para resolver os diversos problemas das cidades, o Big Data surge como uma alternativa aos recursos tradicionais. Como dito por Emmanuel Letouzé (2012), economista de desenvolvimento senior:

“It is time for the development community and policymakers around the world to recognise and seize this historical opportunity to address twenty-first century challenges, including the effects of global volatility, climate change, and demographic shifts, with twenty-first century tools”<sup>3</sup> (p. 6).

De fato, há uma sensação generalizada de que o mundo se tornou mais volátil nos últimos anos e quem sofre são os mais vulneráveis. As crises na economia, não apenas no Brasil, e a crise de representatividade gerada, entre outros, pela falta de empregos, alta no preço de alimentos e da gasolina, se alastraram pelo mundo devido, em parte, pela interconectividade global. Problemas locais

---

<sup>3</sup> Tradução livre da autora: “É tempo para a comunidade que trata sobre desenvolvimento e aqueles que fazem políticas ao redor do mundo reconhecer e tomar essa oportunidade histórica para responder a desafios do século XXI, incluindo os efeitos da volatilidade global, mudanças climáticas e movimentos demográficos, com ferramentas do século XXI”.

geram impactos muito mais longe do que ocorria anteriormente.

Para solucionar essas crises, governos locais não podem mais tratar apenas dos resultados sem se preocupar com as causas. Prevenir os danos ou mantê-los em um nível mínimo é muito mais barato em longo prazo. Para tanto, os meios tradicionais de coleta e utilização de dados, embora tenham sua importância e deem importantes informações, são muito custosos e incipientes, não conseguindo prever soluções adequadas pois, ainda que a causa seja conhecida, não se dispõem de meios de saber ou distinguir quais os grupos afetados, onde, quando e o grau do dano causado.

Enquanto o setor público ainda lida com dados tradicionais, o setor privado vem continuamente sendo bem sucedido na utilização do Big Data na tomada de decisões<sup>4</sup>. No entanto, o setor público e a comunidade internacional vêm acordando para as possibilidades de utilização do Big Data para fins de auxílio na tomada de decisões em políticas públicas, além de diversos fins humanitários.

“It is the double recognition of the promise of the data revolution and the need for better, faster information in an age of growing global volatility that led the leaders of the G20 and the UN Secretary-General to call for the establishment of the Global Pulse initiative (in the wake of the on-going Global Economic Crisis), with the aim of developing of a new approach to ‘social impact monitoring’ and behavioral analysis by building on new sources of data and new analytical tools”<sup>5</sup> (Letouzé, 2012, p. 13).

Através de poderosos algoritmos os dados são extraídos das mais diversas fontes e analisados de forma a se encontrar correlações, padrões. Uma vez treinados (com os dados fornecidos), os algoritmos podem ser utilizados no auxílio de previsões que podem ser utilizadas para encontrar padrões ou ainda anomalias e divergências das tendências esperadas.

No entanto, o tipo de análise utilizado varia dependendo do objetivo que se pretende alcançar. Ela pode ser classificada em Prescritiva, Diagnóstica, Descritiva e Preditiva. Como se poderia esperar, a análise preditiva é a que mais chama a atenção do setor público diante da possibilidade de se utilizar

---

<sup>4</sup> Tem-se como exemplo a pesquisa feita pelo professor do MIT, Erik Brynjolfsson, que constatou diferenças significativas na utilização de dados – uma diferença de 5% na produtividade, considerada como uma vantagem decisiva – percebida por empresas que utilizam “*data-driven decision-making processes*” em face daquelas que continuam a se utilizar principalmente em experiência e intuição (Lohr, 2011).

<sup>5</sup> Tradução livre da autora: “É o duplo reconhecimento da promessa da revolução de dados e a necessidade por informações melhores e mais rápidas na era de aumento da volatilidade mundial que levou os líderes do G20 e a Secretaria Geral das Nações Unidas a requerer pelo estabelecimento da iniciativa Global Pulse (respondendo à crise econômica global em andamento), com o objetivo de desenvolver uma nova visão para o monitoramento de impactos sociais e análise de comportamento através da construção de novas fontes de dados e novas ferramentas de análise”.

da mineração de dados históricos para traçar tendências futuras ou possibilidades futuras.

Um exemplo dessa utilização é o *Google Flu Trends*, lançado em 2008, que se baseia na análise de pesquisas no Google sobre sintomas de gripe. Já há pesquisas indicando que “[...] because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms [...]”, seria possível “[...] accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day” (Ginsberg, Mohebbi, Patel, Brammer, Smolinski, & Brilliant, 2008, p. 1013).

Sendo assim, seria possível a utilização dos dados de pesquisas no Google para detectar epidemias de influenza em áreas com uma alta quantidade de pessoas com acesso à internet. De forma geral, os dados *online* vêm sendo utilizados como parte de vigilância de síndromes de forma a prevenir e combater surtos antes que afetem um grande número de pessoas. O *Google Dengue Trends* funcionava exatamente dessa forma<sup>6</sup>.

Outro caso de utilização do Big Data de forma preditiva foi a captação de informações dadas pelas pessoas após o terremoto no Haiti. Foi instalado um sistema de centralização de mensagens que permitia que fossem mandadas mensagens de texto com informações de pessoas presas em prédios afetados. “According to Ushahidi’s<sup>7</sup> Patrick Meier these results were evidence of the system’s ability to predict, with surprisingly high accuracy and statistical significance, the location and extent of structural damage post-earthquake.”.

Sendo assim, percebe-se que o Big Data tem o potencial de diminuir as falhas humanas e demora na produção, coleta e análise das informações de forma a otimizar a tomada de decisões quando se trata de políticas públicas. Ainda assim, deve-se ter ciência de que a utilização do Big Data, bem como dos algoritmos para sua análise, está longe de ser a solução para todos os problemas da humanidade. Ao contrário, eles próprios apresentam diversas dificuldades na sua utilização, além de colocarem em risco diversos direitos dos cidadãos, como a privacidade e a igualdade.

---

<sup>6</sup> Dados disponíveis em [https://www.google.com/publicdata/explore?ds=z3bsqef7ki44ac\\_&hl=en&dl=en#!ctype=l&strail=false&bcs=d&nselm=h&met\\_y=flu\\_index&scale\\_y=lin&ind\\_y=false&rdim=country&idim=country:BR&ifdim=country&hl=en\\_US&dl=en&ind=false](https://www.google.com/publicdata/explore?ds=z3bsqef7ki44ac_&hl=en&dl=en#!ctype=l&strail=false&bcs=d&nselm=h&met_y=flu_index&scale_y=lin&ind_y=false&rdim=country&idim=country:BR&ifdim=country&hl=en_US&dl=en&ind=false).

<sup>7</sup> “Ushahidi é uma companhia sem fins lucrativos que foi desenvolvida para mapear informações de violência no Quênia após as eleições de 2007. Ela se especializou em desenvolver softwares abertos e gratuitos para a coleta de informações, visualização e mapeamento interativo”. Disponível em <https://www.usahidi.com>. Tradução livre da autora.

## RISCOS NA UTILIZAÇÃO DO BIG DATA

Quando se fala em utilização do Big Data para uma análise preditiva logo se pensa em sua utilização para fins de combate à violência, de fato um dos maiores problemas atuais em diversas cidades. No entanto, como bem observam Barrocas e Selbst (2016), dentre diversos outros pesquisadores, a utilização de Big Data e algoritmos na tomada de decisões não livra o resultado dos preconceitos e discriminações humanas. Não se trata, nem se deve ter tal expectativa, de uma decisão imparcial.

Como em regra a discriminação não é intencional, o resultado acaba por ser uma distorção velada sob a ilusão de imparcialidade. Além disso, como o mecanismo que leva a essas decisões distorcidas se encontra protegido dentro de uma caixa preta<sup>8</sup> a injustiça acaba por se tornar difícil de identificar.

## COMO O BIG DATA DISCRIMINA

Antes de adentrar nos casos de discriminação, é importante analisar como esta de fato ocorre. A mineração de dados desde a sua coleta até a apresentação de um resultado pode se utilizar de cinco mecanismos que podem levar à distorção e a definição do problema, treinamento dos dados, seleção de dados, utilização de *proxies* e *masking*.

“[...] defining the target variable, labeling and collecting the training data, using feature selection, and making decisions on the basis of the resulting model. Each of these steps creates possibilities for a final result that has a disproportionately adverse impact on protected classes, whether by specifying the problem to be solved in ways that affect classes differently, failing to recognize or address statistical biases, reproducing past prejudice, or considering an insufficiently rich set of factors. Even in situations where data miners are extremely careful, they can still effect discriminatory results with models that, quite unintentionally, pick out proxy variables for protected classes”<sup>9</sup> (Barrocas & Selbst, 2016, p. 671).

<sup>8</sup> O termo *Black Box* foi cunhado por Frank Pasquale, juiz e pesquisador norte-americano, e é utilizado em seu livro “The Black Box Society: The Secret Algorithms that Control Money and Information”.

<sup>9</sup> Tradução livre da autora: “[...] definindo o alvo variável, nomeando e coletando os dados de treinamento, usando a seleção de dados e tomando decisões com base nos modelos resultantes. Cada um desses passos cria a possibilidade de um resultado final que tenha um impacto desproporcional em grupos protegidos, seja por especificar o problema a ser resolvido de maneira a afetar grupos de forma distinta, falhando em reconhecer ou endereçar distorções estatísticas, reproduzindo preconceitos históricos, ou considerando uma gama insuficiente de fatores. Mesmo em situações onde os mineradores de dados sejam extremamente cautelosos, ainda há a possibilidade de surgirem resultados discriminatórios com os modelos que, ainda que de forma não intencional, se utilizem de proxies como variáveis em casos de grupos protegidos”.

É importante considerar que a mineração de dados se trata de uma forma de análise estatística, o que gera sempre alguma forma de discriminação. A própria finalidade do Big Data é justamente prover uma base racional em cima da qual se poderá atribuir a um indivíduo qualidades características de um determinado grupo de forma a melhor decidir sobre esse indivíduo ou grupo.

Ainda sim, a análise de dados tem o potencial de prejudicar, de forma indevida ou excessiva, grupos considerados como vulneráveis ao colocá-los sistematicamente em uma relativa desvantagem. Diferente das demais formas, mais subjetivas, de tomada de decisão, as distorções geradas pela mineração de dados não é ligada diretamente a preconceitos humanos, sejam eles conscientes ou inconscientes, e é justamente nesse ponto que reside seu real perigo.

## CASO DA POLÍCIA PREDITIVA

Diz-se comumente que a arte imita a realidade, mas há casos, como no exemplo do filme *Minority's Report*, em que o ser humano tenta prever o futuro. Embora o momento atual não seja exatamente como o do filme, pode-se considerar que chega bem próximo. Já é real a utilização desses modelos matemáticos no processamento de dados na prevenção de crimes. Cidades como Memphis e Nova Iorque já os utilizam há alguns anos. Recentemente foi criado na Argentina o Observatório Nacional de Big Data que pretende, dentre outros objetivos, realizar análises de dados para previsão de eventos futuros de qualquer natureza. Além disso, a implementação desses sistemas vem sendo discutida em países como a França e o Reino Unido.

Em Memphis há a iniciativa *Blue Crush (Criminal Reduction Utilizing Statistical History)*, onde foram coletadas todas as ocorrências policiais. Estes dados são registrados por aparelhos utilizados pelos próprios policiais e coletados por um software preditivo da IBM que, em conjunto com um serviço de mapeamento da empresa ESRI, cria mapas com os possíveis locais de novos crimes. Se utilizando desses dados, a polícia direciona o patrulhamento de forma a trazer maior precisão e poupar recursos financeiros.

A polícia de Chicago, por sua vez, se utiliza de uma tecnologia denominada *ShotSpotter*, que é capaz de registrar sons de tiros, além de sua localização aproximada. Recentemente, esse sistema foi integrado à tecnologia que utiliza algoritmos preditivos que identificam padrões e tendências a partir

dos dados do *ShotSpotter*. Há ainda a integração com o sistema *HunchLab*, que ajuda a polícia a tomar decisões de acordo com a análise massiva das informações sobre detenções, chamadas 911, atividades de gangues e outros dados relevantes.

Com esses sistemas integrados, a cidade de Chicago conseguiu reduzir a os tiroteios em 49% e 66%, em uma análise feita entre fevereiro e março de 2017, nos dois distritos onde esta foi implementada.

No entanto, o caso mais famoso certamente é a utilização do sistema PredPol, desenvolvido pelo Departamento de Polícia de Los Angeles em parceria com a Universidade da Califórnia. Tal sistema tem como utilizar os dados das forças policiais na identificação dos problemas relacionados à criminalidade e suas soluções, fornecendo ajudas na gestão de recursos e escolha de estratégias de forma a prevenir fatos futuros.

Da mesma forma que nos casos anteriores, o PredPol é utilizado para geoposicionar policiais em local e hora corretos para prevenir um possível crime.

No intuito de não gerar discriminações, o sistema é *color blind*, ou seja, não se utiliza de dados relacionados à raça ou etnia de indivíduos. A predição se baseia em dados que buscam responder a três perguntas: o que (tipo de crime), onde (localização) e quando (data e hora) – não quem.

No entanto, ainda que tivesse a melhor das intenções o sistema ainda gera discriminações, como apontado por O’Neil (2016, p. 87):

“When Police set up their PredPol system they have a choice. They can focus exclusively on so-called Part 1 crimes. These are the violent crimes, including homicide, arson, and assault, which are usually reported to them. But they can also broaden the focus by including Part 2 crimes, including vagrancy, aggressive panhandling, and selling and consuming small quantities of drugs. Many of these “nuisance crimes would go unrecorded if a cop weren’t there to see them”<sup>10</sup>.

Como a própria autora bem afirma, a maioria desses crimes menores ocorre principalmente em áreas

---

<sup>10</sup> Tradução livre da autora: “Quando a polícia instala seu sistema Predpol, eles têm uma escolha. Eles podem focar exclusivamente em crimes denominados como Parte 1. Esses são os crimes violentos, estando aí incluídos homicídio, incêndio criminoso e roubo, que são normalmente reportados aos policiais. Mas eles também podem expandir o foco de forma a incluir crimes denominados como Parte 2, dentre os quais estão a gandaia, pedir esmola de forma agressiva e a venda e consumo de pequenas quantidades de drogas. Muitos desses crimes, denominados como meros incômodos, iriam permanecer como não registrados caso não houvesse um policial no local para vê-los”.

mais pobres de uma cidade. Ao incluí-los no sistema preditivo, a análise é comprometida e os policiais tendem a ser deslocados para essas áreas onde eles têm maior probabilidade de prender um maior número de pessoas, ainda mais considerando que um modelo matemático indicou que naquela localidade são maiores as chances de ocorrer um crime.

Sendo assim, há um risco de violação de diversos direitos, dentre eles a vistoria de pessoas e casas sem um devido mandado judicial ou qualquer prova ou indício de um crime. Na verdade, o sistema acaba não por prever crimes, mas sim prever para onde a polícia efetivamente irá. Além das possíveis violações à privacidade e à justiça, há uma constante vigilância de determinados grupo sociais, quer eles tenham cometido algum crime ou não.

Caso a polícia efetivamente encontre o crime, ainda que de menor potencialidade lesiva, esses dados voltam para o sistema algorítmico, que volta a mandar os policiais para essas mesmas áreas.

“This creates a pernicious feedback loop. The policing itself spawns new data, which justifies more policing. And our prisons fill up with hundreds of thousands of people found guilty of victimless crimes. Most of them come from impoverished neighborhoods, and most are Black or Hispanic. So even IF the model is color blind, the result of it is anything but. In our largely segregated cities, geography is a highly effective proxy for race”<sup>11</sup> (O’Neil, 2016, p. 87).

Deve-se considerar, neste momento, dois pontos: i) que esses crimes de menor lesividade também são praticados por pessoas de maior poder aquisitivo, ainda que em menor escala; e ii) que esses indivíduos em geral praticam outros tipos de crimes ainda mais lesivos para a sociedade em geral. No entanto, esses fatos não aparecem no sistema preditivo de forma que não haverá o policiamento devido.

Sendo assim, o sistema acaba por criminalizar, na verdade, a própria pobreza. E o faz de forma a parecer não apenas científico, mas justo.

---

<sup>11</sup> Tradução livre da autora: “Isso cria um pernicioso *loop* de *feedbacks*. O próprio policiamento cria novos dados que justificam mais policiamento. E nossas prisões ficam lotadas com centenas de milhares de pessoas condenadas por crimes sem vítimas, ou seja, sem grande lesividade. A maioria delas vem de bairros empobrecidos e são, em sua maioria, negros ou latinos. Então, mesmo que o modelo seja indiferente à cor, o resultado é tudo menos isso. Nas nossas cidades segregadas a geografia é um dado aproximado muito eficiente para raça”.

## CASO BRASILEIRO

Em 2014, o governo do Estado de São Paulo firmou uma parceria com a Microsoft para a utilização do sistema denominado Detecta, que se utiliza do Big Data para análise dos dados da polícia e de radares e câmeras de vigilância. Há atualmente 2.661 câmeras ligadas aos batalhões, delegacias e, inclusive, *tablets* das viaturas. Associado ao sistema Omega da Polícia Civil de São Paulo, que coleta informações de cadastros de registros civis, criminal, de armas, de veículos roubados e furtados, da junta comercial, disque-denúncia, delegacia eletrônica e DETRAN, além de mapas e sistema de identificação biométrica, os dados analisados podem ser utilizados para investigar, mapear e combater a criminalidade no Estado.

As polícias judiciárias estaduais, de modo geral, ainda estão se modernizando, coletando dados úteis e desenvolvendo técnicas de análises diagnósticas e até descritivas para investigações. Nada próximo ao que ocorre nas cidades americanas. Poucas possuem processos de análise preditiva como fonte investigativa para estratégia de policiamento e prevenção de crimes.

A Polícia Federal, por sua vez, já está bem a frente nesse quesito, tendo uma gama de softwares, como o Nudetective<sup>12</sup> e o EspiaMule<sup>13</sup>, que permitem o correlacionamento de dados de diversas fontes. Apesar de ainda se utilizar principalmente de descritivas e diagnósticas, ela já possui a capacidade de implementar também unidades preditivas.

No entanto, foi com a operação Lava Jato que a polícia federal efetivamente adentrou na era do Big Data ao criar um banco de dados unificados, além de procedimentos analíticos, que foi denominado como “o Big Data para o combate à corrupção”.

Diante do extenso banco de dados e dos mecanismos já existentes para a sua análise, é possível, futuramente, a adoção de uma análise preditiva, caso os dados sejam alimentados constantemente, de forma a se traçar perfis e tendências para o combate à fraude e outros crimes.

---

<sup>12</sup> O Nudetective foi desenvolvido pelos peritos criminais Pedro Eleutério e Matheus Polastro e é utilizado na detecção de pornografia infantil por meio de análises de imagens, nomes, *hashs* e vídeos. O Nudetective possui princípios que podem ser explorados para expansão conceitual e técnica instituindo uma análise forense direcionada para uma coleta massiva de dados em tempo real.

<sup>13</sup> O EspiaMule realiza pesquisas e coletas de informações de usuários da rede Emule, catalogando endereços e criando um mapa da distribuição de imagens de pornografia infantil, o que atualmente é realizado também na rede P2P *Utorrent* e técnicas são aplicadas também na *Deep Web*.

No entanto, há que se estar atento às possíveis violações a direitos fundamentais. Ainda que não haja distorções nos resultados preditivos, como no caso do sistema PredPol, o que não se pode afirmar visto que não se sabe com clareza como os algoritmos funcionam ou quais são os dados utilizados, há uma flagrante violação da privacidade dos cidadãos.

## CONCLUSÕES

Considerando as informações expostas pode-se constatar que o Big Data tem um enorme potencial quando se trata de políticas públicas, podendo ser utilizado no auxílio na tomada de decisões em diversas questões prementes nas cidades.

No entanto, como verificado, problemas no próprio algoritmo ou em relação aos dados permite que resultados distorcidos, de forma consciente ou inconsciente, passem muitas vezes despercebidos, prejudicando de maneira excessiva grupos considerados como vulneráveis.

Ainda que sejam percebidas distorções, estas são difíceis de serem provadas em âmbito judicial, tendo em vista a opacidade dos sistemas ou dos dados utilizados. Tal falta de transparência é defendida pelas empresas proprietárias dos algoritmos, que alegam sua proteção sob o manto de segredo industrial. Abrir a caixa preta, segundo essas empresas, geraria o roubo dos seus segredos por concorrentes, o que geraria um dano irreparável.

No caso do sistema PredPol, só se pôde ter conhecimento da discriminação gerada quando, em 2015, a empresa aceitou, diante da pressão social, abrir seu sistema para análise de pesquisadores.

Há uma necessidade premente de regulação não apenas da coleta de dados, mas também da utilização desses algoritmos preditivos. Trata-se de informações sendo utilizadas pelo poder público para o bem-estar social de forma que o benefício pode ser mais abrangente, mas os prejuízos também são.

## REFERÊNCIAS

Barrocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, 104(1). Recuperado em 30 de outubro de 2017 de [http://Dennys%20Game/Downloads/Big%20Data%20disparate%20im](http://Dennys%20Game/Downloads/Big%20Data%20disparate%20impact%20-%20Solon%20Barrocas%20(1).pdf)

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2008). Detecting Influenza Epidemics Using Search Engine Query Data. *Nature*, 457(1), 1012-1014. Recuperado em 30 de outubro de 2017 de [http://static.googleusercontent.com/external\\_content/untrusted\\_dlcp/research.google.com/en/us/archiv](http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en/us/archive/papers/detecting-influenza-epidemics.pdf)

Kohn, S. (2013, junho 03). Big Data é o novo petróleo, afirma executiva da IBM. *Olhar Digital*. Recuperado em 30 de outubro de 2017 de <https://olhardigital.com.br/noticia/big-data-e-o-novo-petroleo,-afirma-executiva-da-ibm/34986>.

Letouzé, E. (2012). Big Data for Development: Challenges & Opportunities. *United Nations Global Pulse*. Recuperado em 30 de outubro de 2017 de <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>.

Lohr, S. (2011, abril 24). When There's No Such Thing as Too Much Information. *The New York Times*.

O'Neil, C. (2016). *Weapons of Math Destruction*. Nova Iorque: Crown.

Pasquale, F. (2016). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge: Harvard University Press.

The Economist (2017, maio 06). *The world's most valuable resource is no longer oil, but data*. Recuperado em 30 de outubro de 2017 de <https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource>.