# DECONSTRUCTING DATA MINING: PROTECTING PRIVACY AND CIVIL LIBERTIES IN AUTOMATED DECISION-MAKING

Lindsey Barrett[*]

CITE AS: 1 GEO. L. TECH. REV. 153 (2016)

https://perma.cc/LB3S-Y3K3

INTRODUCTION ........................................ 153

UNDERSTANDING THE BASICS ........................................ 155

PRIVACY IMPLICATIONS IN DATA MINING ........................................ 156

DISCRIMINATION IN DATA MINING ........................................ 157

REDUCING DISCRIMINATION IN ALGORITHMIC CONSTRUCTION AND DATA MINING ........................................ 158

## INTRODUCTION

"Big Data" is the bogeyman of the information age: powerful, and as ill-defined as it is abstractly threatening. Broadly, it encompasses "technology that maximize[s] computational power and algorithmic accuracy";[1] "types of analyses that draw on a range of tools to clean and compare data";[2] and the underlying belief in the correlation between the size of the data set, and its ability to produce increasingly accurate and nuanced insights.[3] Put another way, "'Big data' [is] the amassing of huge amounts of statistical information on social and economic trends and human behavior."[4] The belief in the prescient value of big data has led to widespread collection of information on citizens and consumers in both the public and private sectors, though that distinction has

[*] Managing Editor, GLTR; Georgetown Law, J.D. expected 2017; Duke University, B.A. 2014. © 2016, Lindsey Barrett. This piece is adapted from a memorandum I wrote as a summer clerk at the Electronic Privacy Information Center. A description of EPIC's work on algorithmic transparency, and a compilation of related resources, can be found at https://epic.org/algorithmic-transparency/.

[1] Meg Leta Ambrose, *Lessons from the Avalanche of Numbers: Big Data in Historical Perspective*, 11 I/S: J.L. & POL'Y FOR INFO. SOC'Y 201 (2015); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 96 (2014).

[2] Crawford & Schultz, *supra* note 1, at 96.

[3] *Id.*

[4] *Id.*

become increasingly permeable.[5] Data brokers, companies that create and sell detailed profiles of consumers for profit, sell their products to private and public entities alike, and often do not have data quality control clauses in the contracts governing those interactions.[6] These profiles also often refer directly or indirectly to sensitive attributes, such as race, gender, age, and socioeconomic status.[7]

This brave new world of big data is no longer new. But the mechanics of the algorithms relying on that data, and the process by which decisions are made using that information, merits a sharpened focus. Algorithmic decision-making is increasingly replacing existing practices in both the public and private sector, making an understanding of the technical construction of those algorithms increasingly crucial. This is all the more true for processes in which the consumer or citizen does not have a voice, and the logic behind the decision is fundamentally opaque.[8] It is difficult, if not impossible, for that consumer or citizen to challenge an adverse decision made about her when the basis for the decision is unavailable. In the private sector, automated predictions are used to calculate loan rates, credit scores, insurance risk, employment evaluations, and in hiring searches.[9] In the public sector,[10] they are being used for risk prediction

[5] Victoria Schwartz, *Overcoming the Public-Private Divide in Privacy Analogies*, 67 HASTINGS L.J. 143, 189 (2015), at 149 n.31 (overview of literature examining the degradation of the public-private sector divide).

[6] FED. TRADE COMM'N, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY, 16 (2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf (noting the contracts between data brokers and their sources rarely address the accuracy of the provided information).

[7] *Id.*

[8] Jenna Burrell, *How The Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, BIG DATA & SOC'Y, Jan. 2016, at 5, http://bds.sagepub.com/content/spbds/3/1/2053951715622512.full.pdf.

[9] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671, (2016); Rick Swedloff, *Risk Classification's Big Data (r)evolution*, 21 CONN. INS. L.J. 339 (2015); Frank Pasquale, *We're Being Stigmatized by 'Big Data Scores We Don't Even Know About*, LA TIMES, (Jan. 15, 2016), http://www.latimes.com/opinion/op-ed/la-oe-0116-pasquale-reputation-repair-digital-history-20150116-story.html.

[10] *See generally*, Julie E. Cohen, *The Regulatory State in the Information Age*, 17 THEORETICAL INQUIRIES IN L. 2 , (forthcoming 2016 ), http://ssrn.com/abstract=2714072; *Id*, at 13 ("In an era when decisionmaking is mediated comprehensively by so-called "big data," regulators will have to contend with the methods by which regulated decisions are reached — i.e., with the algorithm as an instrumentality for conducting (regulated) activity").

in law enforcement,[11] as well as for sentencing,[12] and to calculate benefits.[13] Further, there is a pervasive and misguided belief in the inherent neutrality of algorithmic decision-making by virtue of its empiricism. But data is not inherently neutral, and neither are the algorithms that process it. Each is the product of the beliefs, fallibilities, and biases of the person who created them. If those fallibilities are unaccounted for, algorithms will simply replicate the pre-existing inequalities encoded in their intake data and structure. This memorandum will provide an overview on the basics of algorithms and data mining, and explore how automated decision-making can unintentionally reveal sensitive information, or unintentionally base their predictions on protected traits, implicating individual privacy and civil liberties.

## UNDERSTANDING THE BASICS

To understand how the particular features of an algorithm can violate individual privacy, or lead to discriminatory outcomes, it is necessary to understand the discrete steps of how algorithms work. An algorithm can be defined as "simply a series of steps undertaken in order to solve a particular problem or accomplish a defined outcome."[14] In the context of big data, that means a computational process that takes input data and creates an output based on a rule.[15] A machine-learning algorithm involves two distinct processes: a classifier algorithm, and a learner algorithm.[16] A classifier algorithm performs a mathematical function on a given set of input data, and creates a category based on the relationships between different properties ('features' of the data) as an output. An example would be a classifying algorithm that takes a list of emails with multiple features, such as sender, time sent, or presence of an attachment, and sorts them by sender ("from Bob" or "not from Bob"). The learner algorithm will establish the relationships between a set of features in

[11] Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 EMORY L.J. 259, 273 (2012).

[12] Sonja B. Starr, *Evidence-Based Sentencing And The Scientific Rationalization Of Discrimination*, 66 STAN L. REV. 803 (2014) (discussing the use of risk prediction algorithms in sentencing and bail determinations).

[13] Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008); Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1510 (2013) (discussing the use of predictive models in IRS audit selections).

[14] Nick Diakopoulos, *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*, TOW CTR. FOR DIGITAL JOURNALISM, 4, (2013), http://www.nickdiakopoulos.com/wp-content/uploads/2011/07/Algorithmic-Accountability-Reporting_final.pdf.

[15] *Id.*

[16] Burrell, *supra* note 8, at 3.

training data, and prospectively apply that rule to new inputs.[17] Commonly used machine-learning models include neural networks, decision trees, Naïve Bayes, and logistic regression.[18] The choice of model depends on the particular use, such as an algorithm designed to predict creditworthiness, as opposed to an algorithm designed to predict the likelihood of crime in a given area, and different models can be used separately, or in conjunction with one another.[19] A prioritization algorithm, as the name might suggest, ranks an input by virtue of possession or lack of certain attributes, and is primarily used in processes that assess risk. Examples include recidivism algorithms used by judges in sentencing, or algorithms that assess insurance risk.[20]

## PRIVACY IMPLICATIONS IN DATA MINING

The very value of data analytics lies with is its ability to elicit subtle and insightful relationships between various data features, such as, oddly enough, an increase in Pop-Tart purchases before hurricanes.[21] That seemingly oracular ability to illustrate connections between otherwise random attributes is both what make big data so useful, and what leads to its piercing ability to reveal private information. It can elicit inferences an individual did not want to know, or might not want anyone else to know, such as a medical condition.[22] It can also draw relationships between legally protected and unprotected categories, and base decisions off of those correlations.[23] Even when the information is not legally protected or inherently sensitive, there are concerns that increasingly precise determinations could be used to create inscrutably complex portraits of consumers, in a way that could further diminish consumer control.[24] Privacy violations and discriminatory outcomes are a predictable consequence of data analytics' ability to elucidate unexpected information. While distinct concepts, privacy and civil rights often overlap when the private information is deeply

[17] *Id.* at 5.

[18] *Id.*

[19] *Id.* at 5.

[20] Starr, *supra* note 12, at 825.

[21] Andrej Zwitter, *Big Data Ethics*, BIG DATA & SOC'Y, Nov. 2014, at 4, http://bds.sagepub.com/content/spbds/1/2/2053951714559253.full.pdf.

[22] Crawford & Schultz, *supra* note 1, at 97 (discussing how a health condition can be inferred from data on consumer habits).

[23] Barocas & Selbst, *supra* note 9.

[24] Solon Barocas, *Data Mining and the Discourse on Discrimination*, CTR. FOR INFO. TECH. POL'Y (2014), 2, https://dataethics.github.io/proceedings/DataMiningandtheDiscourseOnDiscrimination.pdf.

connected to a fundamental right, or a protected attribute, such as political affiliation, immigration status, or a disability.

## DISCRIMINATION IN DATA MINING

Algorithms can be intrinsically (and unintentionally) discriminatory through the population of data selected, how the algorithm functions, and the data itself. For example, when the training data for a predictive policing algorithm assigning the probability of crime to an area uses crime statistics from police stops in 1956 Chattanooga, the algorithm will learn—and replicate— a correlation between arrest rates and race. Data does not simply occur; it is created, and will reflect the flaws of its creator, as will any rule predicated on the relationships between various attributes in that data.[25]As a matter of technique, machine learning is also less accurate, and thus roughly less effective, for minority groups. There is proportionately less data available for majority groups by definition, and correlations that may be correct for the majority may be completely incorrect for the minority.[26] In an excellent piece illustrating the fallacy of inherently neutral data mining, Moritz Hardt uses the example of a machine learning algorithm distinguishing between real and fake names.[27] A short and common name might be real in one culture, and fake in another; if the classifier discerns a negative correlation between real names and complex or long ones, it will be inaccurate in applying that rule to minority groups. [28] Certain attributes can also serve as proxies for sensitive attributes, such as race, or socioeconomic status. Uber, for example, was accused of redlining by directing drivers away from majority-black neighborhoods.[29] Inference of membership in a protected class; statistical bias skewing the function of the algorithm; and faulty inferences based on mistaken or acontextual data can each serve the render the results of an algorithm discriminatory, or violate an individual's privacy.[30]

[25] JONATHAN STRAY, THE CURIOUS JOURNALIST'S GUIDE TO DATA 7, (2016) https://www.gitbook.com/book/towcenter/curious-journalist-s-guide-to-data/details.

[26] Moritz Hardt, *How Big Data Is Unfair*, MEDIUM, (September 6, 2014), https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de#.asxzmuhfg.

[27] *Id.*

[28] *Id.*

[29]Jennifer Stark & Nicholas Diakopoulos, *Uber Seems to Offer Better Service In Areas With More White People. That Raises Some Tough Questions,* WASH. POST (Mar. 1, 2016), https://www.washingtonpost.com/news/wonk/wp/2016/03/10/uber-seems-to-offer-better-service-in-areas-with-more-white-people-that-raises-some-tough-questions/.

[30] Barocas, *supra* note 24.

## REDUCING DISCRIMINATION IN ALGORITHMIC CONSTRUCTION AND DATA MINING

The problems big data poses for privacy and civil rights are manifold and complex. Though the work ahead is considerable, technologists and legal scholars have begun exploring relevant techniques to better guard against discrimination and protect individual privacy. Computer scientists in public policy like Latanya Sweeney,[31] Cynthia Dwork,[32] Helen Nissenbaum[33] and Moritz Hardt[34] have shed light on the fallacy of inherently neutral data mining through research on techniques to combat discrimination, and protect privacy. These technical approaches include both discrimination-blind, as well as discrimination-aware data mining,[35] privacy-aware data mining,[36] and differential privacy.[37] Legal scholars have also begun to delve deeply into how the mechanics of data mining, and the myth of its assumed neutrality, often undermines the assumptions predicating existing laws.[38] The Federal Trade Commission's Big Data report summarized relevant questions for engineers working with large data sets and trying to ascertain the risk of privacy violations

[31] Latanya Sweeney, *Discrimination in Online Ad Delivery*, DATA PRIVACY LAB (2013), http://dataprivacylab.org/projects/onlineads/1071-1.pdf.

[32] Cynthia Dwork, *Differential Privacy: A Survey of Results,* THEORY & APPLICATIONS OF MODELS OF COMPUTATION 1, 19, (2008), https://www.researchgate.net/profile/Minzhu_Xie2/publication/220908334_A_Practical_Parameterized_Algorithm_for_the_Individual_Haplotyping_Problem_MLF/links/0deec5328063473edc000000.pdf#page=12.

[33] Helen Nissenbaum, *A Contextual Approach to Privacy Online*, DAEDALUS 4, 32-48 (2011), http://ssrn.com/abstract=2567042.

[34] Ilias Diakonikolas, Moritz Hardt, & Ludwig Schmidt, *Differentially Private Learning Of Structured Discrete Distributions, in* ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 2566-2574, (2015).

[35] Cynthia Dwork et. al., *Fairness Through Awareness*, ARXIV, (Nov. 2011), https://arxiv.org/abs/1104.3913. (Proposing a framework for fair classification comprising (1), a (hypothetical) task-specific metric for determining the degree to which individuals are similar with respect to the classification task at hand; and (2), an algorithm for maximizing utility subject to the fairness constraint, such that similar individuals are treated similarly).

[36] Sara Hajian & Josep Domingo-Ferrer, *A Methodology for Direct and Indirect Discrimination Prevention in Data Mining*, IEEE TRANSACTIONS ON KNOWLEDGE & DATA ENG'G 25, no. 7 (May 21, 2013) (Proposing a pre-processing discrimination prevention framework to prevent direct discrimination, indirect discrimination, or both, with the objective of a fair tradeoff between discrimination removal and data quality).

[37] Moritz Hardt, Katrina Ligett, & Frank McSherry, *A Simple and Practical Algorithm for Differentially Private Data Release,* http://www.moritzhardt.com/papers/mwem.pdf.

[38] Citron, *supra* note 13; Barocas & Selbst, *supra* note 9.

or inherent discrimination, such as whether a relevant model accounts for biases, and closely the dataset mirrors the population being measured. [39]

Ultimately, preliminary research is exactly that—preliminary. It does not answer all the tough questions raised by the use of big data, and how automated decision-making challenges existing legal frameworks designed to protect privacy and civil liberties. While understanding the mechanics of algorithmic decision-making is fundamentally necessary to prevent violations of privacy and civil liberties from simply being ignored, it is only the first step towards preventing them. At the very least, it is a start.

[39] FED. TRADE COMM'N,, BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION? (2016), https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf; ("How representative is your data set?...Does your data model account for biases?...How accurate are your predictions based on big data?...Does your reliance on big data raise ethical or fairness concerns?").