Predicting Human Behavior

by Alex Rosenblat, Tamara Kneese, and danah boyd

A workshop primer produced for:
The Social, Cultural & Ethical Dimensions of "Big Data"
March 17, 2014 - New York, NY
http://www.datasociety.net/initiatives/2014-0317/

Brief Description

Countless highly accurate predictions can be made from trace data, with varying degrees of personal or societal consequence (e.g., search engines predict hospital admission, gaming companies can predict compulsive gambling problems, government agencies predict criminal activity). Predicting human behavior can be both hugely beneficial and deeply problematic depending on the context. What kinds of predictive privacy harms are emerging? And what are the implications for systems of oversight and due process protections? For example, what are the implications for employment, health care and policing when predictive models are involved? How should varied organizations address what they can predict?

Detailed Topic Description:

There is a strong belief in the potential of "big data" to solve anything from pressing social problems to longstanding business challenges. Some herald its arrival as a way to anticipate and prevent crime or customer attrition. By mining data to both identify new patterns in how people behave, and to predict certain outcomes based on those patterns, predictive analytics promises to translate data into superior and actionable knowledge. Data pioneers collect and combine information on individuals and their activities (e.g., purchasing or credit histories, location data, and search engine queries) with the expectation that these will reveal patterns that hold useful predictive value. Ideally, these details will offer consistent clues about a range of behaviors, from individual preferences to large-scale social trends. From these assemblages of information, an intimate and potentially accurate picture of someone may emerge. From this detailed portrait, further inferences may still be drawn.

Predicting human behavior can be both hugely beneficial and deeply problematic. Drawing on different types of (sometimes sensitive) data to learn something new about human behavior presents exciting opportunities, but also serious policy, legal, and technical challenges. In the best cases, predictive analytics can be used to track movement patterns after natural disasters like landslides, thus helping emergency personnel determine where to go first in a crisis. Prediction seems more ethically ambiguous, however, if children's brain scans and genetic information are used to preemptively pinpoint potential criminals. While such predictive measures may allow for positive early intervention and treatment, helping at-risk

children receive physical activity, better nutrition, and mental health care, they may unnecessarily and unfairly label kids as social deviants before they've done anything wrong. How can we weigh some of the anticipated benefits of predictive analytics against the concerns that arise when drawing and acting on certain inferences?

This tension is obvious in the case of health care. Health analytics may take into account a wide variety of data, including the demographics, personal habits, moods, and other trackable characteristics of patients as well as other traditional diagnostic criteria. But they may also rely on seemingly non-health related data to infer health conditions or predict health outcomes. For example, using information about patients' living arrangements or other aspects of their social lives as variables, health care analytics may be able to predict which patients are most likely to be hospitalized in the near future, granting healthcare providers the opportunity to intervene in advance with the possibility of saving lives and reducing costs. While this research can potentially provide valuable insights into optimizing health care, especially in terms of uncovering the social factors undergirding health outcomes and empowering individuals to become more proactive about their own health, patients can easily be de-anonymized with their demographic information (zip code, birthdate, and gender) alone, nevermind 'bigger' data on them. However, anonymization or de-anonymization is not necessarily the issue here: instead, it is important to consider that analytics may take away an individual's capacity to choose with whom they share their health information. Balancing the potential benefits with the potential consequences is a key challenge in predictive analytics.

Sometimes, it may be worth trading anonymity or risking de-anonymization for insight. In order to predict outbreaks of diseases, organizations may require non-anonymized data to track the likelihood of outbreaks, based on factors that can be gleaned from specific, and identifiable, instances in the course of a patient's disease treatment. Yet, even when data is intended for certain kinds of societally beneficial predictive work, it is sometimes redirected to new, less societally acceptable uses. Recently in the UK, the National Health Service was barred from selling patient information to insurance companies or commercial entities. The NHS had gathered information on patients in order to track national health trends, but this information was then sold for commercial purposes. Insurance companies who received patient data then adjusted premium prices accordingly for groups of customers based on more detailed information about illnesses for specific groups. Given how health information can be repurposed to multiple uses, some more socially acceptable than others, what kinds of barriers to health data access should exist, and when should these barriers be lifted? Even making supposedly anonymous data available can lead to problems. In Washington, patient-level health data is sold for $50 per record, and by matching individual records to news stories containing the word 'hospitalization' in 2011, news reporters were able to accurately identify individuals in 35 out of 81 cases by name.

Can the same privacy tools be used for all health data, or for other domains of data? For instance, pharmaceutical companies use seemingly innocuous personal details about people, like their credit card history or whether they have cats or premium cable TV or drive an American-made car, as highly accurate predictive indicators of which medical conditions they

might have. They use these predictions to locate and solicit potential participants in clinical trials without gaining access to their health records directly. This brings into question the notion that data is sensitive only when it deals with health explicitly.

While there are sometimes mechanisms for individuals to learn what data has been collected about them, the current data ecosystem presents a new challenge for individuals or other entities seeking to learn what is known about them or what inferences have been drawn about them. Even when such information is accessible, it may be hard for individuals to understand how particular criteria play into (sometimes rather consequential) decision-making. Persuasive technologies can present information in such a way that predicting human behavior is partially about molding or changing it, such as by making some products, adverts, or sources more accessible than others, but the subtle differences between prediction and influence can be challenging to distinguish between. The features that make a person or a group of people amenable to specific kinds of influence may defy intuitions. Algorithms may determine what kinds of content individuals see, making assumptions based on previous behaviors, demographic information, or other personal data. This can affect purchasing decisions if only certain offers are presented and may affect employment, housing choices, or educational decisions in the same vein. If algorithms lead to discrimination, this fact could be obfuscated unless every part of the data collection and analysis process is made transparent. How can the dangers and benefits of transparency be weighed? When should transparency be part of a data collection and analysis system's protocol and when should it be obfuscated?

Various organizations use predictive variables from found data to assess who is likely to do what. For example, Chicago's police department used an algorithm inspired by sociological studies on criminal behavior that is designed to help locate who might be more likely to commit violent crimes. Those on the so-called "heat list" come under police supervision, whether or not they are under investigation for any past or present criminal activity. How do such predictions affect the criteria that figure in the police's "reasonable suspicion," and what is the impact on freedom of association? And when does targeted surveillance of this sort amount to a form of discrimination?

The power of predictions is accompanied by the possibilities and potential for harm, although it can be challenging to hold an algorithm, its creators, or its implementers accountable for false positives, or offensive outcomes. In a marketing context, misidentifying consumers' preferences might result in problematic adverts, but most mis-targeted ads are harmless. In contrast, what happens when someone is subject to an erroneous inference about her health? Even here, there's a difference between the dangers that such errors pose in the contexts of medical practice, insurance, or quantified self projects. The implications of erroneous predictions differ further still when we're talking about criminal justice.

In certain contexts and under certain conditions, even accurate predictions can arouse fears. For example, an advert that appears to offer alternative treatments for breast cancer, drawing on multiple, de-contextualized cues (e.g., book purchases from Amazon, search queries, fitness trackers, genetic tests like 23andMe, etc.) can be jarring, especially when one hasn't publicly disclosed a negative health status, or even told friends and family. What kinds

of predictive privacy harms emerge from the aggregation of found data? Where do the opportunities and challenges lie? How will poor use of data get in the way of societal benefits? What are the implications for systems of oversight?

Case Study 1: Human Resources

Bank of America hired Sociometric Solutions to track its employees' behavior so that it could increase productivity. Sensors were attached to employees' badges that monitored employees' movements in the workplace, with whom they speak, and their tone of voice. After analyzing their findings, Bank of America determined that workers who take breaks together are more productive after their breaks because they use the time to get stressful items off their chests. Sensor activity indicated that workers voices sounded less stressed as a result of group breaks. This project can be interpreted in multiple ways. On the one hand, it's fantastic that the company found ways to increase productivity by reducing stress among its employees. On the other hand, this kind of analytics ushers in a new type of relationship between employer and employee, undermining employee agency and freedom even at break time. While new analytics tools may benefit employers and even employees, what are employers' ethical responsibilities to their employees in using these tools? How do these tools affect the power relations between employees and employers?

All too often, when analytics are employed, there's an assumption that the data should be able to speak for itself. Yet, if 'the data' indicates that there is a rational basis for decisions that result in unfavorable outcomes for specific social groups, how should companies or industries address that? For instance, some companies, particularly those who self-insure to provide health coverage to their employees, encourage strong health and wellness habits in their employees, with the aim of reducing how much employees have to use their health insurance. Should employers use wellness program data to determine the prices of their health coverage premiums of their employees or the attractiveness of job applicants? Can that data lead to negative inferences about other aspects of an employee's performance? Will employees with relatively lower incomes, greater familial obligations, or longer commuting times be effectively penalized compared to their colleagues if they either cannot afford the money or time for healthier lifestyles? Generally, if the outcome of predictive analytics is discriminatory, even if the process for arriving at that outcome is not motivated by prejudice, what other criteria should inform such decision-making?

Case Study 2: Education and Financial Aid

University admissions departments try to find students who are extremely likely to succeed, and who will accept offers of admission for the least amount of financial inducement by using a variety of predictive variables. The result is often perversely counter-intuitive to the idea behind financial aid, which is ostensibly to allow lower-income students similar access to higher education as more privileged students. For instance, ProPublica examined data from Newman University, and found that students from lower-income backgrounds paid several

thousands more than higher-income students overall. Universities rank higher if they distribute aid to a larger portion of their incoming class, so taking a full tuition scholarship, like $20,000, for one lower-income pupil and dividing it into four scholarships of $5,000 to distribute to students from higher-income backgrounds achieves ranking-points while also luring students who are on the fence or who can be swayed by a small amount of aid to attend.

There is very little transparency from universities about the kinds of criteria used to offer financial aid, but universities seem to use a wide variety of information to determine the likelihood that a student will select them over a competitor. Universities may only make offers to students they anticipate will choose them by analyzing a range of information, from comprehensive financial profiles of their families, or how long they stay on the phone with an admissions officer, to tracking when a student clicked on an email communication or checked his application status. Even the wording of scholarship categories often obscures the significance of what little criteria the universities are willing to reveal in how they make decisions about admissions more generally, or scholarships specifically. (Need-based scholarships also include merit considerations.)

Although financial aid is often publicly understood to be a tool for creating more equitable access to education, the stark reality of the higher ed business is that universities have limited resources and are looking to use them to maximize a variety of important variables, from the quality of their student bodies to the reputation that they garner through third-party rankings. The more that they can use tools to predict outcomes, the better they can optimize their decisions. But is their approach fair? It certainly isn't meritocratic, but university admissions never were, even if the public thinks otherwise. How should universities balance their use of these tools?

Case Study 3: Predictive Policing

Recent research indicates that a person's risk of being involved in gun violence, either as a victim or a perpetrator, depends on her social ties with others that have a history of violence. In other words, your risk of being shot is directly and significantly related to the number and strength of your social ties to someone who has been shot or who has been arrested for violent behavior. Leveraging this insight, the Chicago police department has adopted predictive analytics to anticipate who might be more at-risk of gun violence. They offer educational interventions to those at high risk, in much the same way that health awareness campaigns are used to mitigate the spread of disease. These uses of predictive analytics to prevent people from being subjected to violence can be quite beneficial, but people worry, and with good reason, when these same techniques are used in other ways. The potential for distrust in how these tools are used is vast, and it is challenging to figure out if transparency or other tools might be useful for building or maintaining trust in the institutions that use predictive analytics, and the purposes to which they use them.

Historical crime analysis often produces "heat maps," showing physical spaces that are more likely to experience criminal activity. New techniques, like those employed by the Chicago police department, can now produce a "heat list," where specific individuals are identified as being particularly dangerous, even if they have not previously committed a crime. By combining heat maps with heat lists and drawing on geo-location data from cell phone towers, police departments may be able to predict where crimes are likely occur based on the movements of people, particularly 'suspicious' people, into unusual areas. If police have information that can help them locate areas or people who pose higher risks for criminal activity, how should they handle that data and how should they approach its use?

In some senses, predictive policing has long been underway. Law enforcement officers track known criminals and more frequently patrol certain regions within a city. They also scrutinize certain populations or classes of people more intensely, raising serious questions about civil rights issues. And yet, as the availability and diversity of data increases, there is a shift from targeting classes of people or geographic regions to targeting specific individuals and more precise locations. What does it mean for law enforcement to use data other than official records of criminal activity? Is there a difference between using already-collected police data for predictive analytics, versus using data collected from non-policing access points like social media? Is one data source more legitimate than another? Does the outcome (like reduced crime, if that's the case) justify the means?

Questions to Consider

- What are the major social, cultural, and ethical tensions that emerge because of predictive analytics? What needs to be better understood to address these?
- What conflicting values and tradeoffs are at stake? How do we understand relevant actors, stakeholders, and "camps"?
- How are the opportunities and challenges of predictive analytics different in different domains (e.g., criminal justice vs. healthcare vs. marketing)?
- What are salient case studies that highlight the tensions, tradeoffs, and issues?
- Who should be responsible for addressing different issues that emerge because of predictive analytics? What is the role of the government? Of data providers? Of existing tools? Of educational institutions? Of media?
- Who should serve as a data caretaker? What is the role of the government in supporting, regulating, protecting data caretakers?
- How do we account for bias and error and their potentially discriminatory effects?
- Who is able to challenge algorithmic systems? What is the role of algorithmic transparency? What is the role of reverse engineering algorithms?
- How can we determine what it means to be algorithmically accountable? What would be a reasonable way for people to appeal the veracity or availability of the data published on them by algorithmically-generated results?

- When and where should access to information be curtailed because of the potential for misinterpretation or abuse? Who should get to decide when information should be curtailed?