CENTER FOR DATA INNOVATION

April 21, 2014

Mr. Donald S. Clark
Federal Trade Commission
Office of the Secretary
Room H-113 (Annex J)
600 Pennsylvania Avenue, NW
Washington, DC 20580

RE: Spring Privacy Series: Alternative Scoring Products, Project No. P145401

Dear Mr. Clark,

On behalf of the Center for Data Innovation (www.datainnovation.org), I am pleased to submit these comments in response to the Federal Trade Commission's (FTC) request for public comment on predictive analytics products.[1] The Center for Data Innovation is a non-profit, non-partisan, Washington-DC based think tank that formulates and promotes pragmatic public policies designed to enable data-driven innovation in the public and private sectors, create new economic opportunities, and improve quality of life. The Center is affiliated with the Information Technology and Innovation Foundation.

In these comments we argue that predictive analytics, i.e., the use of data and algorithms to make probabilistic statements about unknown events, are a valuable tool for organizations (businesses, governments, and non-profits) and individuals. Used in a variety of fields, such as finance, health care, and marketing, they are often more accurate than human judgment alone, and they benefit consumers in a variety of ways, such as by personalizing experiences, saving time, and even increasing privacy in some cases. While the technology is new, for the most part, the issues raised by predictive analytics are not. Even before the era of big data, businesses would routinely make predictions about the likelihood of future events, albeit with less precision than they can now. Since existing laws protect consumers from discriminatory practices, even those resulting from the use of predictive analytics, and given the enormous potential benefits of predictive analytics as it improves over time, the FTC should avoid unnecessarily regulating this technology or the underlying data. Instead, the FTC should work with businesses,

[1] "Spring Privacy Series: Alternative Scoring Products," Federal Trade Commission, accessed April 16, 2014, http://www.ftc.gov/news-events/events-calendar/2014/03/spring-privacy-series-alternative-scoring-products.

consumers, and other stakeholders to identify unaddressed harmful uses of predictive analytics and narrowly restrict those uses.

WHAT IS PREDICTIVE ANALYTICS?

Predictive analytics refers to an array of statistical and computational techniques that use data to estimate the likelihood of an unknown event occurring in a particular timeframe. For example, predictive analytics may be used to predict whether it will rain today, whether a machine part will fail during the next month, or whether a county will go “red” or “blue” in the next election. Using predictive models, computers can process vast quantities of data at speeds that far exceed human capacity, and in many cases, produce predictions with accuracy exceeding that of expert opinion.[2] These models generate predictive scores about individuals, things, and events. An online music service may create predictive scores for how likely a listener is to enjoy a particular song, a health care provider may create predictive scores for how likely a patient is to forget to take his medicine, and a business may create predictive scores for how likely an employee is to leave the company.

Making predictions, of course, is not new. History is littered with examples of psychics, astrologers, and other seers who have made claims to being able to forecast future events. And while these forms of pseudoscience have generally been replaced with more scientific methods, there are countless examples of where decisions are still being made based on expert opinion alone rather than evidenced-based analysis or where rudimentary predictive models are used in lieu of more advanced ones. However, this is changing. By using better data and better models, predictive analytics are able to reduce the number of errors in decision-making processes and, in the process, help consumers.

HOW ACCURATE ARE PREDICTIVE SCORES?

A predictive score is an estimate of the likelihood of a particular event occurring within a particular timeframe. Such scores are not definitive claims (e.g., “it will rain tomorrow”) in and of themselves, instead they represent an estimate of the probability that an event will occur (e.g., “there is a 60 percent chance of rain tomorrow”). Data scientists who design predictive algorithms can determine the range of likelihoods in which it is appropriate to make a firm prediction and take action, and this range may vary from application to application. For

---

[2] Derrick Harris, “Researchers say AI prescribes better treatment than doctors,” *GigaOm*, February 11, 2013, http://gigaom.com/2013/02/11/researchers-say-ai-prescribes-better-treatment-than-doctors/.

example, if an algorithm predicts that a driver has a one percent chance of getting into a minor car accident in the next year there may be little need for intervention, but if an asteroid has a one percent chance of striking the Earth, then that might be cause for greater alarm.

Predictive scores do not perfectly estimate the likelihood of events. Models are not perfect, and even if an event is very likely to occur, chance can always intervene. The same randomness that fundamentally precludes weather forecasters or poker players from perfectly anticipating the future also affects prediction in other arenas. Moreover, a model may simply be unable to say with certainty that an event has a high or low likelihood of occurring; sometimes events really do have "50-50" chances. However, even if predictive analytics cannot perfectly forecast events, they are still more accurate than predictions derived from gut feelings or other less-precise sources. One famous example comes from the bestselling 2003 book *Moneyball*, in which Oakland Athletics general manager Billy Beane leads his team to success by selecting players according to factors that were more accurate predictors of winning than the rules of thumb and gut feelings employed by other managers of his time.[3] Beane's chosen predictors did not come close to a 100 percent success rate, but they worked better than those used by his competitors.

Regulators should be careful not to note that a predictive score's current accuracy is not necessarily indicative of its potential future performance. Predictive analytics are based on models that are designed to improve over time as more data is gathered. For these reasons, the accuracy of a predictive score, *per se*, is not a useful criterion for determining its utility.

HOW ARE PREDICTIVE SCORES USED?

Credit scores, which predict the level of risk in extending credit to an individual, are the canonical example of predictive scores. However, predictive scoring is used in fields including health care, education, insurance, and even online dating.

In health care, predictive scores are pervasive. They include well-established scores like life expectancy and more modern inventions such as an individual's genetic predisposition to disease, likelihood of hospital readmission, and likelihood of taking medicine.[4] These scores allow doctors to tailor treatments according to individual patients' characteristics, improving

---

[3] Michael Lewis. *Moneyball* (New York: W.W. Norton & Company Inc., 2003).

[4] Alex Guazzelli, "Predictive analytics in healthcare," *IBM DeveloperWorks*, November 29, 2011, https://www.ibm.com/developerworks/library/ba-ind-PMML3/.

outcomes by dispensing with one-size-fits-all dosages and lowering overall costs by reducing readmissions.

In education, standardized test scores are often used by schools as a predictor of students' likelihood of performing well in subsequent grades or education levels.[5] These scores allow educators to prioritize the scarce resources of supplementary tutoring or other remedial programs to the students who most need help. In recent years, higher education officials have begun to use predictive scores on students' likelihood of succeeding in particular courses of graduating college. For example, Wichita State University implemented a predictive analytics program to reduce the dropout rate for its students and better identify high-quality applicants. The university found that predictive scores were a significantly more reliable indicator of student success than recommendations made by human reviewers.[6]

Insurance providers quantify risk for an extremely broad array of events, including the likelihood that a homeowner will sell, the likelihood a farmer's crop will be wiped out by flood, and the likelihood that a particular stock will increase in price.[7] Predictive scores in the insurance industry can ensure that safer drivers really do get lower rates and that potential home buyers do not get denied a loan for the wrong reasons. Crop insurance company The Climate Corporation creates such precise models based on the likelihood of extreme weather events that it can offer payouts automatically to farmers without requiring them to prove losses. This high level of efficiency allows farmers who grow crops with narrow planting windows to replant before it is too late for that season.[8]

Finally, predictive scores are used in a variety of online recommendation systems, with dating sites modeling how likely it is that two people are romantically compatible, Amazon determining which books a user might like to read, and LinkedIn determining what jobs a user might be

[5] Nathan Kuncel and Sarah Hezlett, "Standardized Tests Predict Graduate Students' Success." *Science* 315 (2007): 1080, http://www.unc.edu/~nielsen/soci708/cdocs/Kuncel_Hezlett_2007_1080.pdf.

[6] Neal Ungerlieder, "Colleges Are Using Big Data To Predict Which Students Will Do Well--Before They Accept Them," *Fast Company*, October 21, 2013, http://www.fastcoexist.com/3019859/futurist-forum/colleges-are-using-big-data-to-predict-which-students-will-do-well-before-the.

[7] Susanne Sclafane. "Use of Predictive Models Widespread in P/C Insurance: Survey," *Insurance Journal*, November 7, 2013, http://www.insurancejournal.com/news/national/2013/11/07/310601.htm.

[8] "Agriculture and algorithms," *The Economist*, November 19, 2012, http://www.economist.com/blogs/schumpeter/2012/11/weather-insurance.

interested in.[9] These scores allow the companies to offer personalized services, saving consumers' time by obviating the need to slog through enormous numbers of listings to find relevant information and compare alternatives.[10]

HOW DOES PREDICTIVE SCORING AFFECT CONSUMERS?

As described previously, predictive scoring can be used to improve consumer welfare by personalizing products and services, including in areas like health care and education. In addition, predictive scoring can often benefit consumers in other ways, such as by increasing privacy, freeing up resources to handle exceptional situations, and rewarding positive behaviors.

First, predictive scores can be used to trigger automated responses, preventing the need to disclose an individual's information to another person. For example, an online store may generate personalized product suggestions automatically, enabling shoppers to avoid revealing their preferences and purchase histories to a storekeeper. Or a health insurance company may send automated reminders to help patients adhere to their doctor's orders, thereby reducing the need for future medical exams. In general, predictive scores reduce the need to communicate sensitive information to people who might divulge it later.

Second, by accurately and efficiently scoring individuals, automating scoring can help make businesses more productive by reducing the need to manually review data about customers. Higher productivity results in savings that are generally passed on to consumers in competitive markets. This shift should also move workers into higher-paying, higher-productivity jobs. For example, instead of a loan officer or admissions officer manually processing thousands of applications where the decision is clear, they can instead spend more time reviewing applications for individuals whose characteristics make them statistical outliers. Automated scoring could free up resources to allow businesses to better address individuals that may be "caught in the machine."

---

[9] Janet Wagner, "Machine Learning and Predictive Analytics Foster Growth," *ProgrammableWeb*. February 21, 2014, http://blog.programmableweb.com/2014/02/21/machine-learning-and-predictive-analytics-key-to-business-growth-and-profitability/; Ryan Neal, "What Is Bright.com? Why LinkedIn Purchased The Job-Matching Website For $120 Million," *International Business Times*, February 6, 2014, http://www.ibtimes.com/what-brightcom-why-linkedin-purchased-job-matching-website-120-million-1553869.

[10] JP Mangalindan, "Amazon's recommendation secret," *CNN Money*, July 30, 2012, http://tech.fortune.cnn.com/2012/07/30/amazon-5/.

Third, predictive analytics makes it easier to reward positive behavior. Individuals may alter their behavior if they know a company is using predictive scoring. While this is sometimes cited as a public policy concern, it is worth noting that many of these changes may be desirable.[11] For example, drivers may better adhere to the speed limit if they know their roadway habits will affect their insurance premiums, or consumers may make more responsible financial decisions in order not to negatively affect their credit rating.

HOW SHOULD THE FTC ADDRESS THE GROWTH OF PREDICTIVE SCORING FOR CONSUMERS?

Imagine that businesses could predict events with perfect accuracy. Some individuals would be better off because their actual risk is less than was previously predicted, while others would be worse off because their actual risk is greater than was previously predicted. Overall, society would be better off because prices would be correctly tied to risk. While we will never have perfect predictions, we are moving towards better accuracy, and this creates more fairness for consumers. The key is for policymakers to reduce the risk of inaccurate data and, in some cases, evenly distribute risk among citizens, such as with health insurance.

Ultimately, predictive analytics offer substantial benefits to consumers in a broad range of fields, and using predictive analytics is superior to other, less-precise, non-computational forms of prediction commonly used today. To ensure consumers continue to benefit from this technology, the FTC should work with the private sector to encourage more widespread use of predictive analytics. To do this, laws and regulations should encourage the use and reuse of data. It is far more efficient to collect information once and use it many times than to collect data again for each new data model. In addition, policymakers should promote the development of tools and techniques to de-identify data, as well as support research on topics like privacy-preserving data mining.

In addition, policymakers should ensure that predictive scores based on positive consumer behaviors do not result in negative feedback to consumers. For example, since it is socially-desirable to have people seek treatment for mental health issues, regulators should make sure this type of information is not used to deny someone housing, employment, or life insurance. Fortunately, many laws, including the Civil Rights Act, Americans with Disabilities Act, the Fair Housing Act, the Fair Credit Reporting, and the Genetic Information Nondiscrimination Act,

---

[11] Daniel Solove, "'I've Got Nothing to Hide' and other Misunderstandings of Privacy," *San Diego Law Review* 44 (2007): 745.

already include restrictions on how data can be used for precisely this reason. To the extent that policymakers find gaps in these laws, they should work to ensure individuals are treated fairly.

Sincerely,

Daniel Castro

Director, Center for Data Innovation
1101 K Street NW, Suite 610
Washington, DC 20005

dcastro@datainnovation.org