The Challenges and Opportunities of Big Data: Reforming State-Based Insurance Regulation in the 21st Century

Federal Advisory Committee on Insurance

Birny Birnbaum
Center for Economic Justice

January 5, 2017

The Center for Economic Justice

CEJ is a non-profit consumer advocacy organization dedicated to representing the interests of low-income and minority consumers as a class on economic justice issues. Most of our work is before administrative agencies on insurance, financial services and utility issues.

***

On the Web: www.cej-online.org

Why CEJ Works on Insurance Issues

Insurance Products Are Essential Financial Security Tools for Individual and Community Economic Development:

CEJ Works to Ensure Fair Access and Fair Prices for These Essential Products and Services, particularly for Low- and Moderate-Income Consumers.

Insurance is the Primary Institution to Promote Loss Prevention and Mitigation, Resiliency and Sustainability:

CEJ Works to Ensure Insurance Institutions Maximize Their Role in Efforts to Reduce Loss of Life and Property from Catastrophic Events and Promote Resiliency and Sustainability of Individuals, Businesses and Communities.

Big Data Defined

Insurers' use of Big Data has transformed the way they do marketing, pricing and claims settlement. Big Data means:

- Massive databases of information about (millions) of individual consumers
- Associated data mining and predictive analytics applied to those data
- Scoring models produced from these analytics.

The scoring models generated by data mining and predictive analytics are algorithms. Algorithms are lines of computer code that rapidly execute decisions based on rules set by programmers or, in the case of machine learning, generated from statistical correlations in massive datasets. With machine learning, the models change automatically.

What’s So Big About Big Data?

1. There has been a revolution in insurance pricing, marketing and claims settlement resulting from insurers’ use of Big Data -- massive databases of new insurance and non-insurance, personal consumer information with associated data mining and predictive analytics and scoring.

2. Insurers’ use of Big Data has huge potential to benefit consumers and insurers by transforming the insurer-consumer relationship and by discovering new insights into loss mitigation.

3. Insurers’ use of Big Data has huge implications for fairness and affordability of insurance and for regulators’ ability to keep up with the changes and protect consumers from unfair practices

4. Big Data has massively increased the market power of insurers versus consumers and versus regulators.

5. Market forces alone – “free-market competition” – cannot and will not protect consumers from unfair insurer practices. So-called “innovation” without some consumer protection and public policy guardrails will lead to unfair outcomes.

6. Oversight and limited regulatory intervention can promote more competitive markets and faster adoption of innovative technologies that benefits consumers and fulfill public policy goals.

Barocas and Selbst: *Big Data’s Disparate Impact*

“In contrast to those traditional forms of data analysis that simply return records or summary statistics in response to a specific query, data mining attempts to locate statistical relationships in a dataset. In particular, it automates the process of discovering useful patterns, revealing regularities upon which subsequent decision-making can rely. The accumulated set of discovered relationships is commonly called a “model,” and these models can be employed to automate the process of classifying entities or activities of interest, estimating the value of unobserved variables, or predicting future outcomes.

“These all involve attempts to determine the status or likely outcome of cases under consideration based solely on access to *correlated* data. Data mining helps identify cases of spam and fraud and anticipate default and poor health by treating these states and outcomes as a function of some other set of observed characteristics.

In particular, by exposing so-called “machine learning” algorithms to examples of the cases of interest (previously identified instances of fraud, spam, default, and poor health), the algorithm “learns” which related attributes or activities can serve as potential proxies for those qualities or outcomes of interest. In the machine learning and data mining literature, these states or outcomes of interest are known as “target variables.”

“The proper specification of the target variable is frequently not obvious, and it is the data miner’s task to define it. In doing so, data miners must translate some amorphous problem into a question that can be expressed in more formal terms that computers can parse.

Through this necessarily subjective process of translation, though, data miners may unintentionally parse the problem and define the target variable in such a way that protected classes happen to be subject to systematically less favorable determinations.“

Examples of Insurer Big Data Algorithms

Pricing:

- Price Optimization/Demand Models
- Customer Value Scores,
- Telematics,
- Credit Scores,
- Criminal History Scores,
- Vehicle Scores,
- FireLine Rating

Claims:

- Fraud Scores,
- Severity Scores

Example: Claim Fraud Scores, Claim Severity Scores

LexisNexis Claim Tools

“LexisNexis (LN) seeks to provide a Single Point of Entry for delivering all of information directly back into a carrier’s system whether from a marketing standpoint, underwriting process or especially the claims part.

“LN has over 10,000 data sources that feed into its infrastructure each month and has contributed information from the industry.

“Claims Data Fill” – deliver data and analytics directly into claims system in the claims process regarding parties, vehicles and carrier information. Used to verify information provided to insurers and provide indicators beyond the data to identify whether a social security number is an indicator of fraud or whether an address provided is a good address. Has an analytic component at first notice of loss and throughout the claim, constantly monitoring the claim looking for fraudulent activities. Real time data verification and enhancement with fraud scoring and attributes

LexisNexis Claim Tools (con’t)

“Example, insured calls in, rear-ended, all I got was license plate:

“Claims Data Fill takes that license plate, reach out to DMV to get vehicle registration to get VIN number, we have policy database and get the carrier and policy information, take the registered owner, go out to public records, pull back their address, date of birth, telephone number, social security, wrap that into a package and put it back into our system, 88% of the time done in less than 5 seconds.

“Take minimum information provided at first notice of loss, provide a fraud score at the initial notice of loss. Daily monitoring of claim every time new information comes in, able to run various scores: fraud scores, severity score.”

Big Data Algorithms Can Reflect and Perpetuate Historical Inequities

Barocas and Selbst: *Big Data's Disparate Impact*

Advocates of algorithmic techniques like data mining argue that they eliminate human biases from the decision-making process. But an algorithm is only as good as the data it works with. Data mining can inherit the prejudices of prior decision-makers or reflect the widespread biases that persist in society at large. Often, the "patterns" it discovers are simply preexisting societal patterns of inequality and exclusion. Unthinking reliance on data mining can deny members of vulnerable groups full participation in society.

New York Times, August 10, 2015: Algorithms and Bias: Q. and A. With Cynthia Dwork

Algorithms have become one of the most powerful arbiters in our lives. They make decisions about the news we read, the jobs we get, the people we meet, the schools we attend and the ads we see. Yet there is growing evidence that algorithms and other types of software can discriminate. The people who write them incorporate their biases, and algorithms often learn from human behavior, so they reflect the biases we hold.

Q: Some people have argued that algorithms eliminate discrimination because they make decisions based on data, free of human bias. Others say algorithms reflect and perpetuate human biases. What do you think?

A: Algorithms do not automatically eliminate bias. . . .Historical biases in the . . .data will be learned by the algorithm, and past discrimination will lead to future discrimination.

Fairness means that similar people are treated similarly. ***A true understanding of who should be considered similar for a particular classification task requires knowledge of sensitive attributes, and removing those attributes from consideration can introduce unfairness and harm utility.***

Q: Should computer science education include lessons on how to be aware of these issues and the various approaches to addressing them?

A: Absolutely! First, students should learn that design choices in algorithms embody value judgments and therefore bias the way systems operate. They should also learn that these things are subtle: For example, designing an algorithm for targeted advertising that is gender neutral is more complicated than simply ensuring that gender is ignored. They need to understand that classification rules obtained by machine learning are not immune from bias, especially when historical data incorporates bias.

White House Report on Big Data

For all of these reasons, the civil rights community is concerned that such algorithmic decisions raise the specter of “redlining” in the digital economy—the potential to discriminate against the most vulnerable classes of our society under the guise of neutral algorithms. . . . .But the ability to segment the population and to stratify consumer experiences so seamlessly as to be almost undetectable demands greater review, especially when it comes to the practice of differential pricing and other potentially discriminatory practices. It will also be important to examine how algorithmically-driven decisions might exacerbate existing socio-economic disparities beyond the pricing of goods and services, including in education and workforce settings

Big Data Defined

Insurers' use of Big Data has transformed the way they do marketing, pricing and claims settlement. Big Data means:

- Massive databases of information about (millions) of individual consumers
- Associated data mining and predictive analytics applied to those data
- Scoring models produced from these analytics.

The scoring models generated by data mining and predictive analytics are algorithms. Algorithms are lines of computer code that rapidly execute decisions based on rules set by programmers or, in the case of machine learning, generated from statistical correlations in massive datasets. With machine learning, the models change automatically.

Barocas and Selbst: *Big Data’s Disparate Impact*

“In contrast to those traditional forms of data analysis that simply return records or summary statistics in response to a specific query, data mining attempts to locate statistical relationships in a dataset. In particular, it automates the process of discovering useful patterns, revealing regularities upon which subsequent decision-making can rely. The accumulated set of discovered relationships is commonly called a “model,” and these models can be employed to automate the process of classifying entities or activities of interest, estimating the value of unobserved variables, or predicting future outcomes.

“These all involve attempts to determine the status or likely outcome of cases under consideration based solely on access to *correlated* data. Data mining helps identify cases of spam and fraud and anticipate default and poor health by treating these states and outcomes as a function of some other set of observed characteristics.

In particular, by exposing so-called “machine learning” algorithms to examples of the cases of interest (previously identified instances of fraud, spam, default, and poor health), the algorithm “learns” which related attributes or activities can serve as potential proxies for those qualities or outcomes of interest. In the machine learning and data mining literature, these states or outcomes of interest are known as “target variables.”

"The proper specification of the target variable is frequently not obvious, and it is the data miner's task to define it. In doing so, data miners must translate some amorphous problem into a question that can be expressed in more formal terms that computers can parse.

Through this necessarily subjective process of translation, though, data miners may unintentionally parse the problem and define the target variable in such a way that protected classes happen to be subject to systematically less favorable determinations."

CEJ
What is Big Data
Regulatory Framework for PC – Risk-Based Pricing
Not Excessive, Not Inadequate, Not Unfairly Discrimiatory
Unfair Trade Practices/Unfair Claim Settlement Practices
Advisory Organizations
Old Old, Old and New School
Examplesof Big Data
Examples of Rate Filings
Examples of RiskClassifications

Insurer Deviation from Risk Based Pricing: Modeling of Rates
Price Optimization
Telematics
Catastrophe Modeling
Regulaory Response
Unlicensed Advisory Organizatins
Machine Learning, Disparate Impact
Regulatory Big Data to monitor outcomes/promote competitive markets and facilitate more efficient and effective market analysis and market regulation
Regulatory resources to reflect new needs – see PBR
Revise risk classification regulatory framework

- Prior Approval
- Consistent with public policy goals
    - Risk-based pricing / Avoid Adverse Selection
    - Promote Loss Mitigation / Degree of Consumer Control
    - Transparency to Consumers/ Accuracy & Completeness of Data
    - Risk-Spreading
    - Avoid/Minimize Disparate Impact

Benefits of Monitoring Outcomes
What does monitoring market outcomes mean?
Regulatory Big Data

- Calibrate inputs/processes to outcomes
- Assist financial regulation
- More efficient/effective market regulation
- Information to evaluate insurer and producer practices

- Information to evaluate existing and proposed policies
- Information to meaningfully evaluate affordability and availability and competition issues
- Information to promote competitive markets