

# **Data Science W231: Legal, Policy, and Ethical Considerations for Data Scientists**

Section 1: Raymond Buhr and Anant Srivastava

Final Research Project - Outline

## **Topic**

Benefits, risks and implications of collecting and utilizing data about the customer for modern retail companies.

## **Statement of the Problem**

Retailers have had success increasing sales by extending analysis outside of the core business to the individual customers, sometimes leading to serious questions of personal privacy infringement. What limitations of data collection and investigative research should corporations be held to in order to protect consumers while still allowing businesses to compete with advanced analytical strategies?

## **Introduction**

- The rapid growth in big data technologies has enabled retailers to gather and process tremendous amount of data about their customers in order to improve sales and add value to products. ([Riglian, TechTarget](#))
- Retailers have data about their customers' browsing patterns, geo-coordinates, likes and dislikes, brands of devices used, health conditions and so forth. This power to collect and experiment with customer data puts a huge responsibility on retailers' shoulders and has several legal, ethical and privacy implications.
- In this project, we will investigate different retail formats such as an online retailer, a traditional retailer etc. and explore ethical and legal issues surrounding the data collection and usage.
- We will evaluate these using various ethical and legal frameworks such as Belmont report, Fair Information Practice, HIPAA etc, as well as through the lens of responsible research for science.

## **Background Research**

- Benefits of Analytics
  - How companies are using data science to accrue information and how they transform that information into actions, such as knowing who and when to make a personalized discount for. ([MIT Sloan Management Review](#))
- Possible Harms
  - Data breaches leading to identity theft and financial loss for the consumer, as well as damaged reputation and ruined trust for the retailer. ([Target, Dec. 2013; dozens of other retails such as Kmart, Sony and UPS throughout 2014](#))
- Data collection Initiatives
  - Loyalty cards, store affiliated credit cards, mobile apps and online store accounts all have the ability to track personal information in a semi-transparent

way. (Still need to identify sources for how different retailers use these methods)

- Application of analytics
  - Article with interview of Target statistician Andrew Pole, [“How Companies Learn Your Secrets”](#), discloses how statistics can identify correlations used to predict more than just shopping behaviour, but also key life events that can wind up having great influence on future shopping habits, such as pregnancy.

### **Theoretical Frameworks, to apply to three example cases of retailers**

- [Belmont report](#)
- Respect for Persons - 1) autonomous agents and 2) protection for those who do are not autonomous
  - Informed Consent
    - Terms of service (TOS) are often long, technical and full of legal jargon that most consumers could not reasonably interpret on their own accord. Does the respect for persons condition call for additional protection of consumers due to the complexity of modern TOS and the increasingly common disregard that many consumers now have towards TOS?
    - Do convoluted TOS allow firms to reasonably get away with using the data they can gather to make predictions about the behavior of specific individuals? Are shoppers who use loyalty cards or register for online accounts aware of the level of detail being aggregated in the retailer's database? Blanket data collection statements that do not specify the detail or goal of the data collection process should not count as informed consent since there still exists a gap in the symmetry of information and understanding of potential future transactions in favor of the retailer.
    - Would an opt-in approach better allow customers to share the appropriate amount of information for their comfort level?
- Beneficence
  - Loyalty discounts are a way to attract customers to purchase specific items. Some companies, such as CVS, take that approach to the next level by issuing coupons for products and related items predicted to be useful based on recent purchases.
  - Is there an appropriate amount of discount offers for personal care items? Could deals on certain embarrassing items potentially cause harm to the consumer based on the interaction with the store employee receiving and reading the coupons first?
  - Some online retailers track browsing cookies to see what products a consumer has been looking at, even if from a competitor.
  - Do targeted ads based on browsing history interfere with the idea of personal choice?

- Is it more beneficial to the consumer for getting a good deal on a desired product or does the constant reminder of the availability of the item and a sudden deal create an economic transaction that could cause an undue imbalance on the consumer's personal finances. Credit card debt can be a serious problem and the constant monitoring and projecting relevant advertisements may only be maximizing benefits of the retailers without minimizing harm to the consumer.
- [Fair Information Practice](#)
  - Purpose Specification
    - As mentioned with TOS in terms of respect to persons, detailed explanation and justification of the use of data collected about individuals should be required of retailers, including data shared between parent and children companies.
    - Consumers face potential harm when personal identifiable information (PII) is shared between related companies. Even if disclosed in the TOS, shoppers may not assume or realize that the information they are providing to an online retailer is being passed on or even sold. Since the amount of PII necessary to trace back seemingly anonymous online actions will continue to decrease, it will become even more important for people to know exactly what and who will be handling the data released.
  - Data minimization
    - In order to protect consumers from the harm in information sharing, retailers should consider the risks to big data collection. Just because the information can be gathered and may be put to good use, that same data could be also be used to ill effect. Without weighing the costs and benefits, retailers fail to respect their customers. Minimizing the amount of data collected that can be categorized as PII should help offset potential danger from data breaches, as well as provide an internal safeguard when TOS lack clarity in regards to purpose of data collection.
  - Accountability and Auditing
    - There have been several high profile incidents in recent news over violations of PII leaks and possibly unethical uses of consumer data ([Target, Dec. 2013](#); [dozens of other retails such as Kmart, Sony and UPS throughout 2014](#)).
- [On Being a Scientist](#)
  - Ethical guide to research, whether academic science or industrial
  - Authors argue that all research is based on trust
    - Scientists trust that other scientific research was conducted in earnest and that the conclusions made were based on real evidence.
    - Society trusts that scientists experiment to find truth, but also not to experiment with malice intent.

- We argue that society also expects industrial researchers to conduct experiments in an ethical manner, with an objective of finding truth and not creating harm through unfair advantages.

### **Hypothetical and Historical cases that demonstrate the ethical quandaries that consumer driven research on the individual scale create**

- Personalization in an online retail setup
  - In an attempt to personalize information on the buyer screen, a retailer can end up manipulating reviews, prices, product displays etc. which can cause emotional distress and health issues etc. e.g. A manipulated review on a fitness pill can harm the consumer in various ways.
- Data breaches from retailers' systems
  - We would like to explore the impacts of data breaches on customers and the responsibilities of retailer in such events. ([Target Dec 2013](#), [20 different companies in 2014](#))
  - How have retailers handled such events? Are there better ways of protecting data and handling such catastrophes?
- Tracking customers life events
  - Some retailers follow you where you go. Personally, I recently moved from one city to another and within a few days of moving I had received some mail with offers on furniture, curtains, etc. for my new apartment.
  - We would like to explore if companies misuse the power that data has provided them through constant monitoring. Is it fair to be targeted for certain life events when you are more likely to vulnerable to make decisions out of convenience or recent influence, such as moving?

### **Analysis of Specific Retailers**

- Example one: Traditional Retailer, Target
  - Business Model- Mostly Store, some online etc.
  - Customer Interaction Channels- Store/online/Mobile/Paper
  - Data Collection methods- In store, ClickStream, RED Card etc.
  - Data usage and Experimentation.
  - Evaluation.
    - Benefits (For Target and Customer)
      - Highly useful and accurate predictive model
      - Analysis creates edge over competition for a while
      - Experience to apply towards future projects
      - Coupons, Discounts, Targeted Ads, Recommendations.
    - Risks/Costs(For Target and Customer)
      - negative effects on brand image, reputation
      - loss of existing customers out of distrust
      - loss of potential customers out of fear
      - Bad Recommendations

- Bad pricing
  - Influencing the customer into a wrong or harmful decision
- Implications of actions (For Target and Customer)
  - Loss of brand trust due to known privacy infringement
  - Loss of brand trust due to exploited security flaws
  - Emotional, psychological harm from embarrassment of releasing news to friends and relatives without consent (father and teen daughter pregnancy)
  - Financial damage to customers during holidays from stolen debit cards
- Example Two: Online Retailer, Amazon
  - Business Model- Online, Amazon Prime etc.
  - Customer Interaction Channels- Store via web or mobile
  - Data Collection methods- Clickstream, Social Media, Amazon card, 3rd party Apps, etc.
  - Data usage and Experimentation
  - Evaluation
    - Benefits
      - Highly useful and accurate predictive model
      - Discounts/Deals/Coupons, resale market (Used products)
      - Product recommendations, personalization, Usability
    - Risks/Costs
      - Bad Recommendations
      - Bad pricing
      - Influencing the customer into a wrong decision
      - Biasing the reviews.
    - Implications of actions
      - pending further research
- Example Three: To Be Determined after further investigation, desire something prolific but operating in a different space than Amazon or Target

## Conclusion

- Analytical Summary
- Lessons learned and recommendations

## Sources

[1] *The Belmont Report*, Office of the Secretary, Department of Health, Education, and Welfare. Published Apr 17, 1979. Available at

<http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html>

[2] *Fair Information Practice Principles*, National Institute for Standards and Technology.

Available at <http://www.nist.gov/nstic/NSTIC-FIPPs.pdf>

[3] *On Being a Scientist: a guide to responsible conduct in research*, Committee on Science, Engineering, and Public Policy, National Academy of Science, National Academy of Engineering, and Institute of Medicine of the National Academies. Published c2009. Available at <http://oskicat.berkeley.edu/record=b18264853~S1>

[4] *A message from CEO Gregg Steinhafel about Target's payment card issues*, Target Corporation. Published Dec 20, 2013. , Available at <https://corporate.target.com/article/2013/12/important-notice-unauthorized-access-to-payment-ca>

[5] *The Big Data Breaches of 2014*, Bill Hardekopf, Forbes. Published Jan 13, 2015. Available at <http://www.forbes.com/sites/moneybuilder/2015/01/13/the-big-data-breaches-of-2014/>

[6] *'Big data' collection efforts spark an information ethics debate*, Adam Riglian, TechTarget. Published Nov 2012. Available at <http://searchcloudapplications.techtarget.com/feature/Big-data-collection-efforts-spark-an-information-ethics-debate>

[7] LaValle, S., E. Lesser, R. Shockley, S. M. Hopkins, and N. Kruschwitz. "Big Data Analytics and the Path from Insights to Value." MIT Sloan Management Review 52, no. 2 (2011). Available at <http://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/>

[8] *How Companies Learn Your Secrets*, Charles Duhigg, New York Times. Published Feb 16, 2012. Available at [http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?\\_r=0](http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0)