Case Study: Inappropriate Use of Surveys

Introduction

The Cambridge Analytica case, as described by N. Confessore (Confessore, 2018) is presumably the first ever use of personal data on scale to inform a political campaign (build voter profiles, to be exact). I was tasked to consider how this could have happened and why it was used. The particular attention should be drawn to the use of innocuous Facebook surveys, as requested in the case study description. As I would like to keep this case study apolitical and strictly case-focused, I will refrain from naming any specific political party. I will focus on a general use case of such voter profiles and then proceed to discuss the 'how' and the 'what for' questions.

What is a 'voter's profile'?

I would like to start with the definition of the voter's profile. I like to think about it in a similar way as data professionals think about customer personas (Gilliland, 2017). It is a set of attributes which are used to group users into cohorts that are meant to be representative of a population. Such simplification can enhance any marketing (or political campaign) efforts. It makes it easier for a human to visualise and conceptualise an imaginary person (profile) as if they were standing in front of them. This makes it much easier to deliver the content that such a person is likely to want to hear.

The concept of targeted marketing (buyer persona) has been around since 1999 (Zambito, 2010) and has since evolved significantly. Customer profiling (Loshin et al., 2013) can enhance understanding of a company's users (Jansen et al., 2020) in the same way as a voter's profile can enhance a candidate's understanding of the people who might or might not vote for them. It is therefore clear that there is a tremendous value in possessing such a voter's profile. The issue begins where we start looking into how the source data for profiling has been collected.

How was the data accessed?

The case study description suggests (University of Essex, 2023) that the data used by Cambridge Analytica was taken mainly from Facebook surveys. In reality, it has been harvested from Facebook in one of the biggest data leaks in history (Confessore, 2018). Now that we have a clear understanding of where the data comes from, let us consider how that happened. I will leave to cyber security specialists any detailed analysis on how exactly this

much data could have been harvested from secure Facebook servers and I will focus on the surveys mentioned in the case study description (University of Essex, 2023).

As N. Confessore rightly points out (Confessore, 2018): 'You are the product'. In the evolving world and the Internet of Things reality, we share more and more of our data. Often unconsciously, sometimes by mistake but rarely with a pure intention, making informed decisions to supply our information to create profiling for monetary, political or any other gains. If something is for free it is likely that we are the product (Roy, 2018) and more precisely the information about us. I believe the participation in said surveys took place only because they looked innocuous, hence no one suspected that it might be used for informing a political campaign.

Target pregnancy prediction model

Another example of a misuse of surveys and a variety of other data sources about a customer can be found within one of the biggest chain stores in the US: Target. Target has made the headlines when the company became 'aware' of a teen girl being pregnant before her father did (Hill, 2012). This information was predicted by Target's ML model being trained on vast amounts of data and creating customer profiles in return - similar to those described at the beginning of this paper.

It is self-explanatory that for-profit companies exist indeed for profit and therefore are always finding new ways to increase their sales and margins. When it comes to data, however, the ethics question comes to mind immediately, especially in the case of a teen pregnancy described above. We, as a society, have recognised that with growing accessibility of personalised data there is a growing ethical responsibility (Hemker et al., 2021). We shall look forward to seeing how this can be implemented on scale.

Summary

Inappropriate use of surveys is unethical in its definition. If it is inappropriate, then it is almost certainly unethical. It can have profound social and legal implications. In the case of Cambridge Analytica (Confessore, 2018), the misuse has triggered investigations from the highest structures of government in both the US and the UK. The world had seen Mark Zuckerberg (the Facebook CEO) in front of the congress. The 2016 US election was questioned by many (Auchard, 2018). These are only a few of the legal implications such an inappropriate use of surveys can have.

It is not free from social and professional implications either. In Target's case we have learned about customers being 'creeped out' by the algorithm (Hill, 2012), which can negatively impact customer's trust to the brand and to the industry in general, which in turn can have overall negative effects on the economy. Needless to say what social impact can cause an election that is questioned. It can lead to mass protests, sometimes violent (Kishi et al., 2020).

It can lead to periods without a government which again, can negatively impact a country's economy and various other aspects of social life.

It is therefore clear that inappropriate use of surveys and other kinds of data has all sorts of negative implications, which we shall avoid at all costs.

Reference list

Auchard, E. (2018). *Cambridge Analytica CEO claims influence on U.S. election, Facebook questioned*. [online] U.S. Available at: https://www.reuters.com/article/us-facebook-cambridge-analytica-idUSKBN1GW1SG [Accessed 6 Sep. 2023].

Confessore N., (2018). Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. *The New York Times* . https://advance.lexis.com/api/document?collection=news&id=urn:contentItem:5S1M-8HD1-JBG3-63X3-00000-00&context=1519360.

Gilliland, N. (2017). *What are customer personas and why are they important? - Econsultancy*. [online] Econsultancy. Available at: https://econsultancy.com/customer-personas/.

Hemker, S., Herrando, C. and Constantinides, E. (2021). The Transformation of Data Marketing: How an Ethical Lens on Consumer Data Collection Shapes the Future of Marketing. *Sustainability*, [online] 13(20), p.11208. doi:https://doi.org/10.3390/su132011208.

Hill, K. (2012). *How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did*. [online] Forbes. Available at: https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/?sh=16754a686668 [Accessed 6 Sep. 2023].

Jansen, B.J., Salminen, J.O. and Jung, S.-G. (2020). Data-Driven Personas for Enhanced User Understanding: Combining Empathy with Rationality for Better Insights to Analytics. *Data and Information Management*, [online] 4(1), pp.1–17. doi:https://doi.org/10.2478/dim-2020-0005.

Kishi, R. and Jones, S. (2020). *Demonstrations & Political Violence in America: New Data for Summer 2020*. [online] ACLED. Available at: https://acleddata.com/2020/09/03/demonstrations-political-violence-in-america-new-data-for-summer-2020/ [Accessed 6 Sep. 2023].

Loshin, D. and Reifer, A. (2013). *Chapter 8 - Customer Profiling*. [online] ScienceDirect. Available at: https://www.sciencedirect.com/science/article/abs/pii/B9780124105430000081.

Roy, S. (2018). *Facebook: If something is free, 'you' are the product*. [online] TechHQ. Available at: https://techhq.com/2018/04/facebook-if-something-is-free-you-are-the-product/ [Accessed 6 Sep. 2023].

University of Essex (2023). 'Reflective Activity 2', *Research Methods and Professional Practice August 2023 A*. Unit 5.

Zambito, T. (2010). *The Origin of Buyer Personas*. [online] Buyer Persona Insights. Available at: https://www.buyerpersonainsights.com/2010/01/the-origin-of-buyer-personas.html [Accessed 6 Sep. 2023].