University for Business and Technology (UBT)

Course: IT, Ethics, Law and Communication

Cambridge Analytica:
Ethics and voter manipulation

Professor: Dashmir Istrefi

Students: Kushtrim Pacaj,
Kujtim Hyseni,
Kujtim Krasniqi

April 2020

1. Background

Cambridge Analytica was a British based consulting company, which helped many political campaigns in the last few years. They are most known for their involvement in support of Donald Trump's 2016 presidential campaign, Ted Cruz in the Republican presidential primaries, and in the Leave faction of the UK's referendum on whether to leave the European Union.

While it's common for political campaigns to hire consultants such as Cambridge Analytica in order to help their cause by either providing advice, or helping in political messages and ads, the Cambridge Analytica team took it one step further, as we'll show next...
- A couple years ago, a researcher (data scientist) named Aleksandr Kogan created a Facebook app called "thisisyourdigitallife", which was marketed as a personality quiz that users could take and it would assess their personality traits based on their answers. Other than the answers, the app also required users to allow it to access their profile's data (such as likes list, friends' list etc.).
Using that app, Kogan built a very large dataset of Facebook user information, estimated upwards of 87 million users. Then he went ahead and shared the dataset with Cambridge Analytica. It has been reported that the company also gathered data using other 'apps' such as 'Sex Compass' and 'Music Personality', so the number of users they have data on might me even higher.

Now using the data collected, they were able to build a "profile" of the users. This profile allowed some insights on what the users views were politically, for example if they were liberal or conservative. It also allowed them to group people in personality traits, and then tailor and target ads specifically to those people. The idea was to know what people want, and to show them an ad that supports their point of view.
For example if a voter is someone who cares a lot about stopping illegal immigration, you can target an ad telling that once in office, you'll try and halt all illegal immigration. You can also do the opposite for the voter that wants to help immigrants, you can target an ad to them telling that once in office, you shall do everything in your power to help immigrants. This is one reason why micro-targeting is bad and unethical and harmful.. You can promote diametrically opposite views to different people, and obviously once in office, you cannot do both. So you just 'lied' to some voters.

Cambridge Analytica based on the data gathered, it siloed voters in the OCEAN personality traits model, which is an acronym for

1. Openness, - "Do they enjoy new experiences?"

2. Conscientiousness, - “Do they prefer plans and order?”
3. Extraversion, - ”Do they like spending time with others?”
4. Agreeableness, - “Do they put people’s needs before theirs?”
5. Neuroticism, - “Do they tend to worry a lot?”

The idea behind this was to have an understanding of personalities, since according to a presentation by Alexander Nix[2] : “ it is personality that drives behaviour, and behaviour influences how you vote”.
According to a report by Quartz,[1], this is exactly the kind of targeting that Cambridge Analytica excelled and was proud to provide. The article very interestingly shows these practices in action. They show several ads by John Bolton in support of Republican candidate for the Senate ,Thom Tillis of North Carolina. All these ads are targeted at specific people based on their OCEAN score.
For example, for people that were deemed “neurotic” (“neuroticism” is a measurement of how much you tend to worry.), they were shown an ad that plays on those worries by emphasizing the risks to national security from overseas, and the need to have strong leaders.
Or for people they deemed “agreeable” they showed an ad titled “A Safer World for Our Children”. The list goes on, each personality having their own tailored ad.

Ultimately, Cambridge Analytica’s involvement in elections, their targeting and how they got the data on millions of users was on full-on spotlight after a whistleblower came forward with their testimony. Company got chided for breach of trust, for using the user's data without their understanding and consent. In 2018 they filed for insolvency, and the company closed.

2. Analysis - Law and ethics

Every state is responsible for administering the electoral system. As the foundation of functioning democracy and the rule of law, the right to vote de jure must be respected and enforced. Citizens’ rights such as statutory rights—due process, equal representation before the law, the right to appeal, and trial by jury—and constitutional rights like freedom of expression, voting, and non-discrimination are constitutional rules that must be guaranteed.
The State Government takes this as its responsibility and adopts policies to ensure the independent process of voting against fraud, manipulation and falsification.
Governmental legitimacy derives from people’s will. And the will of the people, in turn, emerges from the wills of individual citizens. The ultimate source of governmental legitimacy lies in the political opinions and preferences of the citizens.
These opinions and preferences are, in the form of votes, the input into the democratic system.

Cambridge Analytica, as mentioned above, was a British political consulting firm that combined misappropriation of digital assets, data mining, data brokerage and data analysis with strategic communication during the electoral processes. Scientists question that Cambridge Analytica claims about the effectiveness of its methods of targeting voters. Furthermore, the company closed operations in 2018, with legal proceedings including bankruptcy and members of the SCL Group have been continuing operations under the legal entity Emerdata Limited.
UK big data voter opinion influencer Cambridge Analytics was hired by Donald Trump campaign in the 2016 election. Data was initially gathered a few years ago properly by psychology professor Aleksandr Kogan through Facebook. At the time, Facebook allowed Kogan to collect information from users who downloaded his app, which offered a personality test. Those Facebook users also gave him permission to collect data from their friends, but he wasn't allowed to pass the data along to a third party and what he did was that Kogan passed that data on to SCL Group and Cambridge Analytica, which was working to develop techniques that could be used to influence voters. All this information was used to build psychological profiles of voters. Facebook has suspended Cambridge Analytica and SCL from its social media platform, and asked Cambridge Analytica to agree to an audit in an attempt to show that the data in question was deleted because they breached its rule.

Getting users' Facebook data was one aspect of the legality of Cambridge Analytica's actions and the second was the effectiveness of Facebook advertisements and the ethical questions about them. Cambridge Analytica used unethical practices to collect Facebook data and potentially influence American voters. CA chief data officer, Alex Tayler is recorded separately as saying the firm's analysis was responsible for Trump's Electoral College performance.

In March 2018, it was revealed through undercover footage that Cambridge Analytica used seductive women to entice a rival candidate while secretly videotaping the encounter. The firm also sent impostors who acted like wealthy individuals only to give them bribes.

Christopher Wylie, a former contractor, disputes that Cambridge Analytica destroyed the user data, he is the self-described whistleblower who has shared his story with the press. Cambridge Analytica's goal was to persuade users to vote in favor of their client, which involved displaying messages that would most likely result in a change of behavior. Cambridge Analytica's personality-targeting are tactics that are used now as standard practice in political marketing. Quartz uncovered an international data firm that literally copied Cambridge Analytica's presentation and presented the same methods to clients. While it may seem invasive to predict voters' personality type and target political ads accordingly, there are no US laws that prohibit such practices. So even though this might not have been illegal for now, it is without a question unethical.

Any privacy and data protection legislation should include the following principles, based on the forthcoming “Personal Privacy, Awareness and Control” and should include:

1. Public transparency
2. Disclosure for users,
3. Control and Notification.

The most overarching bill comes from the offices of Senators Richard Blumenthal and Ed Markey, titled CONSENT Act or “Customer Online Notification for Stopping Edge Provider Network Transgressions”, requires FTC to establish privacy protection for customers of online edge providers. Then, The Social Media Privacy protection and Consumer Rights Act of 2018, introduced by Senator Amy Klobuchar, draws similar constraints to the CONSENT Act mentioned before, but adds restrictions on modifications to privacy terms, provisions regarding withdrawal of consent and procedures when a violation of privacy has occurred.

3. Analysis from an ICT professional POV (point-of-view)

We all have read and mentioned hundreds of times that Cambridge Analytica have gathered the data and then sold it to potential clients such as President Trump's electoral campaign team. Around 87 million records were scrapped from Facebook. But, why did it matter and how? Why would somebody pay to know friends of an X voter, pages they like, what singer they like or all this kind of data? How is it all linked to a political campaign and how did it help president Trump gather more votes?

In data science, no matter if you’re going to use a fancy machine learning algorithm or a very simple algorithm, you need a dataset. And in order to have a dataset, you will need to do some things in the old fashioned way.

- Phase 1 - “A personality quiz” which required a facebook login
  This phase included making people go through a 120 questions personality quiz. Persuading people to take such a long survey and filling it correctly is not easy. Some groups of people were more likely to fill in the 120 questions, while some others were harder to convince. Over populating the data with data from only one group (let’s say white women over 30) would lead to uncertainty of the results and predictions. So, Cambridge Analytica started to pay people to fill in the survey. The amount paid for one survey ranged from 2$ to 4$ with the more money being paid to the people who were harder to get. Based on Cambridge Analytica data, african men were the least likely to take a survey, this way earning the most money since the data were overpopulated with

"wealthy white women". This personality quiz was used to profile people in 5 categories using the so-called "Ocean" model of the personality. These categories are:

- Openness - Imagination, feelings, actions, ideas
- Conscientiousness - Competence, self discipline, thoughtfulness, goal driven
- Extraversion - Sociability, assertiveness, emotional expression
- Agreeableness - Cooperative, trustworthy, good natured
- Neuroticism - Tendency towards unstable emotions

This categorization made it easier for the people with these data on their hands to predict their behaviour with a higher reliability but it also helped in evaluating the reliability of the data itself. For example, people who described themselves as "loud" were much more likely to describe themselves as "gregorious" too. So, profiles who have checked both of those options as true, were more trustworthy than the ones who checked only one of them.

- Phase 2 - Collecting the likes, friends list and all data they could collect

  Facebook came into play at the very end. In order to be paid for their survey, users were asked to login with their Facebook account, and this was the time that the application took all the data it could, including likes, friends, etc… From the user experience it was easy, they logged in and they got the payment code without knowing what was really happening in the background. Even though the survey itself didn't ask for real data such as name and location, their Facebook profile did. This made it possible to take all this data and match with a real person who is a real voter. Same kind of data was taken for all of the users' friends.

- Phase 3 - Personality quiz results were paired with facebook data such as likes in order to create a psychological pattern. So, at this phase, you already have the personality quiz and their real name and location. Connecting this with their "like" habits and tendencies, gave the algorithm the opportunity to find out even more.

- Phase 4 - Algorithms combined data with other sources like voters records with hundreds of data points per person and this would make it possible to target people with a very specific kind of advertisement. And all these data were put inside a model. At this point, different machine learning algorithms were used. 253 algorithms were built to make predictions, because each algorithm would have its strong and weak points and this means that 253 predictions were made for each profile. This way the goal was achieved. Cambridge Analytica now had an algorithm that could take facebook likes and do a kind

of “reverse engineer” all the way down to people's personality, habits and preferences. Those 253 predictions for personalities were what distinguished Cambridge Analytica from others. For example, it was made possible to target a neurotic, agreeable Democrat with a very different advertisement than an emotionally stable, introverted Democrat. Even if you will give the same message for three kinds of people, you have the opportunity and the knowledge on how you should talk to them. If you want to emphasize opening new jobs, you take a different approach to different kinds of people. If you’re talking to a conscientious person you could emphasize the opportunity to succeed and the responsibility that it comes with the job, if you’re talking to an open person, you emphasize the opportunity to grow as a person, or if you’re talking to a neurotic person you emphasize the security that this job will give to you to you and to your family.

4. Conclusions and recommendations

What Cambridge Analytica did is considered unethical by most people. But though the way they got the data too was shady (and illegal), the core of what they did, the voter manipulation itself by using micro-targeted ads is sadly not illegal. Though it might not be so for long, since some countries are looking to regulate the online political micro-targeting field.

This whole fiasco would be unthought of 20 or 30 years ago. No one had such detailed data on voters, and no-one had a way to get them easily. But with the advances in ICT, with the development of social media, and billions of users they have, today this kind of data-driven approach to voter manipulation is very much possible.
Possible should not mean that it should be done though. This is a sphere where technology advanced faster than the laws regulating it, and they are left playing catch-up.
And until there are such laws, all that remains is the ethics of doing so.

References

1. https://qz.com/1782348/cambridge-analytica-used-these-5-political-ads-to-target-voters/
2. Cambridge Analytica - The Power of Big Data and Psychographics https://youtu.be/n8Dd5aVXLCc?t=172
3. https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie