Executive Summary

Data Ethics Case Competition Submission

Joon-Ho Son, Kelvin Zhang, Wafia Zia, Maddy Richer

September 2018

Contents

1 Introduction

Innovation in data science has brought major opportunity to the field of public health, but with it a range of complex challenges and considerations have been introduced which must be addressed before action can take place.

Ultimately, our goal is to **inform the government on how to best implement and improve public health schemes based on data we have collected during this project**. We intend to take a data-driven approach in order to provide extensive automated health recommendations based on our findings. Our analysis will be able to provide insight on some of the key problems facing those most vulnerable in society on a range of time and population scales. In this Executive Summary, we will aim to identify and address the ethical and technical challenges, both in and outside of our domain.

2 Status Quo

Historically, longitudinal studies have been limited in the specificity of data that they have collected, and the age and size of government institutions can be a barrier to change; while housing reports are commonplace, such as the annual report published by the GLA (Greater London Authority), these make no attempt to harness emerging technologies such as the *Internet of Things*. It is this chance to bring a unique perspective and innovation in the field that makes this opportunity so exciting.

3 Ethical Values

Ethics and innovation are not mutually exclusive. At every step of our process, we have been guided by a set of core ethical values and principles.

Transparency

We firmly believe that stakeholders have a right to a meaningful explanation for any decision made about them. To qualify as "meaningful", explanations should accomplish the following:

1. Inform the subject as to why a particular decision was made
2. Provide grounds for recourse and contest
3. Ensure understanding of what actions can be taken by the subject to alter a similar decision in the future

Moreover, subjects must consent to and be adequately informed about data collection.

Accountability

A core principle that we believe should apply to any company that deals with the data of individuals is that they are responsible for both the ethical storage and use of the data.

Fairness

Data should be used in a way that treats and benefits all subjects equally - regardless of age, gender, race, or any other protected characteristic.

Purpose

Data should only be collected with a clear purpose as to the benefit that it would provide to stakeholders. Consequently, only the minimum amount of personal data required to achieve a particular goal should be collected.

4 Proposal

4.1 Overview

Our solution focuses on the collection of data in a non-intrusive and primarily passive manner, which can be analysed to determine the health of residents. In our system design, we address how data is collected and analysed and present a discussion on the ethical factors in making our decision.

4.2 Data Collection

In order to maximise the effectiveness of our solution, we first identified common lifestyle factors affecting health, focusing on factors common to occupants of public housing and those from low socioeconomic backgrounds. We finally decided upon factors which could be monitored in non-intrusive manners with devices such as low-cost internet-connected sensors and wearables.

Appliance and Utility Monitoring

We propose the installation of sensors measuring the use of appliances and utilities (i.e. runtime and resource usage), where patterns of usage could relate to health factors. For example, the frequency of shower and washing machine usage can indicate the personal hygiene of a resident, and potentially even mental well-being. By monitoring utilities, either directly (e.g. per kilowatt hour of electricity) or indirectly (e.g. internet-connected smart bulbs), it is also possible to collect data which could correlate to residents' sleeping patterns, another factor of health explained in further detail below.

Wearables

Wearables are an effective way to gain accurate information about health and lifestyle factors such as amount of exercise and heart rate. We propose providing residents a smart watch intended to be worn both day and night, providing large amounts of high quality data such as sleep quality and heart rate at the expense of greater intrusiveness for those unfamiliar to wearables. When worn during exercise, it also allows our platform to gauge the frequency and quality of residents' exercise routines with lower levels of intrusiveness. Resting heart rate is a common indicator of fitness level, and this data combined with information about the subject's lifestyle will help to inform how fitness level can be increased across the population. Recent consumer wearable devices have the capability to perform electrocardiograms in order to detect heart conditions such as atrial fibrillation and sleep apnea - a dangerous and underdiagnosed disorder associated with 33% mortality (Marshall 2008).

General Health Questionnaire

One of the key problems faced by those in social housing is that of mental illnesses and substance abuse. Mental health is a difficult characteristic to measure, as it is highly subjective. Self-reporting can be considered an unreliable way to collect data, but in spite of recent developments we believe that technologies developed to passively monitor the mental health of subjects have not sufficiently matured. One way to monitor the mental health of a population is to ask residents to answer one question each day via a mobile or tablet app; this helps to determine signs of social isolation and drug and alcohol abuse, while also gathering data on the residents' lifestyle. Questions are repeated over time to see if there has been a change in any of the answers.

4.2.1 Harmonisation and Aggregation

There is no reason that data collection needs to be limited to these sensor systems. In fact, we strongly recommend that already available data should be incorporated into any data analyses to be carried out. For example, this could include census data, or data from state-run institutions.

Case Study: Safestats

It has been estimated that 40% of violent crime goes unreported in the UK (Office for

National Statistics 2017). The GLA's Safestats crime data system collates data from hospital A&E departments in an effort to measure and locate unreported crimes. When victims appear at A&E, details of the incident location are recorded. This data is then cleaned, split into relevant fields, matched with location databases, and crucially includes a confidence measure on the location.

Due to data being collected from such a range of sources, it is necessary to implement robust methods of data standardisation. In particular, we suggest the establishment of a separate team responsible for this task, as well as the maintenance of a secure endpoint as an avenue for data sharing. This makes it easy to provide tightly controlled access for different government and third-party organisations to standardised data.

4.2.2 Security and Publishing Control

When handling private and indeed potentially sensitive data, the importance of its proper treatment of cannot be understated. This challenge of data security is not only a technical one, but also ethical as the handling and sharing of data must also be carefully considered.

The importance of de-identification is multi-faceted. First, leaked data about a person's health can be used by employers and insurance firms to discriminate - regardless of the legality of this. Even if we operating under the bold assumption that the data store is secure against outside attackers, one can imagine a scenario where a government employee can access information that could be used unethically. However, perhaps the most compelling reason for us is simply the human right to privacy. We strongly believe that even with consent to use data, effort must be made to preserve a person's privacy to the greatest extent.

In accordance our principles, we recommend sharing and publishing data only in a de-identified state. This goes further than *anonomisation*, as even data that has no direct-identifiers attached to it can be reidentified from quasi-identifiers (for details see Garfinkel 2015). Various levels of de-identification can be applied, depending on the user's authorisation level.

Generating Synthetic Data

A possible solution to this issue is applying modified PCA (Principal Component Analysis) transformations to augment features in such a way that valuable data integrity is retained, but the raw data is obfuscated. In essence, PCA is a statistical procedure in which a dataset is transformed into a lower dimensional dataset while trying to retain as much information as possible. Conventional PCA aims to maximise variance in data, but as this can often reduce discriminating power (for classification) it has been suggested that it may be preferable to instead select components based on the importance of a property in order to maintain statistical attributes (Gal et al. 2014).

4.2.3 Privacy and Consent

The greatest ethical dilemma that has arisen throughout the development of our project is treading the fine line between monitoring and surveillance. For example, we class monitoring social media as invasive, despite the useful indications that it might provide about a person's mental health. Philosopher Michael Lynch offers an enlightening thought experiment in which you imagine an alien with the power to view all of your most intimate thoughts. This invasion ultimately appears violently invasive to the average human being, because without their consent it violates standards of "moral person hood" (Selinger 2014). Even if this alien does nothing with this information and never interacts with any human on earth, it still breaks a moral barrier.

Today we have a clear example of how privacy can be breached by groups with power and technology. The "Prevent Scheme" in the UK monitors possible Muslims at risk of radicalisation, however, instead of providing security for the general public, instead causes fear and unrest among vulnerable groups

in the community (Thomas 2016). In this way our project faces the same problem with a different context: how to monitor a vulnerable group in society without causing unrest, mental health problems, and protecting the dignity, privacy, and right to self-determination of participants.

With this in mind, we have designed the method of data collection in such a way that it will have minimal effects on the lives of subjects, and should not restrict their freedom in any way. The devices will not perform any form of "nudging" in order to reduce their impact on subjects' decisions; however studies such as Oulasvirta et al. (2012) demonstrate that a person's behaviour is often influenced by the knowledge that they are being monitored.

Informed consent is essential in any scientific study involving individuals, its role being to protect and respect subjects' autonomy. During the informed consent process, we will explain how privacy will be maintained, as well as detailing the circumstances under which intervention and the release of results would be necessary. Privacy is a fundamental human right, however we believe breaching confidentiality is sometimes necessary if there is a major net benefit to the person[1], for this reason intervention and identification will only occur if there is a genuine concern for the subject's health.

For consent to be valid, subjects must have *capacity*, defined as the ability to understand information and make decisions based on it. We value fairness, and aim to avoid excluding any person from participating in the study because of their age or health. In fact, given the nature of the aim of our study, a diverse sample is strictly necessary. Nevertheless, vulnerable groups of people such as the elderly and the mentally ill may have reduced autonomy and therefore we will take careful steps to ensure their capacity before allowing their participation in the study.

4.3 Data Use

4.3.1 Analysis Methodology

Our analysis approach can be considered in two separate parts: spotting underlying trends, and detecting anomalies. In any case, a variety of both supervised and unsupervised techniques can be applied.

Anomaly Detection

Anomaly detection is an area in data mining concerned with the identification of items that deviate significantly from the majority of the data. The application of these techniques to this particular project is broad. For instance, anomaly detection has been used to detect unusual patient management in order to identify potential mistreatment (Hauskrecht et al. 2012), but also to detect tumours in medical images (Taboada-Crispi et al. 2009). This is a flexible system that can be used to detect both failing hospitals and patients especially at risk - in which direct intervention may be necessary.

Data Patterns

This is an expansive field of data science, encompassing popular techniques such as k-means clustering, support vector machines, and gradient boosting classifiers. In particular, these algorithms could be applied to evaluate whether there are sub-types of behaviour that lead to certain symptoms by grouping samples into different clusters based on some measured data.

4.3.2 Providing Recommendations

At the core of our proposal is the challenge of how we can provide actionable and above all *useful* items to inform the government. As a solution to this, we present the novel concept of an automated

[1]Truthfully, "net benefit" is a rather ambiguous phrase; in particular, the state deciding "what is best for you" is currently very relevant. Earlier this month, the *European Court of Human Rights* ruled that the UK government's bulk interception regime "violated Article 8 of the European Convention on Human Rights (right to respect for private and family life/communications)" (European Court of Human Rights 2018). Critically, what sets this proposal apart is the collection of explicit consent.

recommendation system that delivers digestible (a property that is increasingly important in this bloated information economy) recommendations based on real-time data in x month intervals. This has the twofold benefit of having items grounded in data instead of speculation, and allowing the government to track the state of public health over time.

In accordance with our principle of transparency, we endeavour to provide meaningful explanations as to why a recommendation was provided. Hence, it is critical to have interpretable evidence that justifies an outcome. There are multiple technical barriers to this. Explaining the functionality of complex decision-making algorithms is difficult, and they are often termed "black boxes" as reference to their opaqueness. This makes it hard to deliver explanations that are of value to data subjects. Moreover, revealing too much information raises concerns about the disclosure of both proprietary software and data that may violate the privacy of other subjects.

As a result, we have adopted the concept of algorithmically generating *counterfactuals*. Wachter, Mittelstadt, and Russell (2017) offer this as the fundamental form of a counterfactual:

> "Score p was returned because variables V had values (v_1, v_2). If V instead had values (v_3, v_4) then score q would have be returned."

This structure is concise and reveals minimal proprietary information, while meeting the characteristics of a meaningful explanation as outlined in our ethical principles. For example:

> "The Freedonia Capital Hospital Pediatric Inensive Care Unit has been classified as *Inadequate* because its number of bed spaces is 181. If it had 214 bed spaces, it would have been classified as *Outstanding*."

It is worth noting that the scope of suggestions could be adjusted based on whether you're discussing an issue on a micro or macro scale, and based on the audience to provide more insight into the data that generated this outcome[2]. In this way, the need to understand the inner-workings of the algorithm is obviated, and instead the explanation is presented in terms of an external dependency[3]. We hope that this approach will contribute to engendering trust with stakeholders, not only limited to our uses, but also play a role in introducing algorithmic decision making to the wider population.

This system also has potential to be adapted into tools that widen access to data and resulting data analysis to specialists from a variety of fields; we think it is important to provide tools that do not restrict data to just data scientists. We envisage that these tools will allow users to visualise and identify relationships and ontologies in data - akin to *Palantir's* existing *Gotham* and *Foundry* applications.

5 Conclusion

In this document we have summarised our proposal and the ethical challenges surrounding it. In the interests of length, aspects such as novel passive mental health monitoring methods, legal considerations, and the technical details of our system architecture as a whole have been omitted from our summary, but provide an opportunity for further discussion. Looking to the future, we would be keen to open a dialogue on the validity of consent given under the incentive of cheaper rent and the issues surrounding preserving protected characteristics in algorithms, but also as to how we can incentivise honest and regular responses to our General Health Questionnaire.

We hope that our data-driven solution brings positive reform for the residents of Freedonia, and will be held as an example of how data can open new avenues for innovation and insight in public health.

[2]One must also consider whether or not controlling this level of information could be used in a selective and biased way, but this is out of the scope of the present discussion.

[3]The nature of this external dependency is important as it would contradict our aim of providing actionable items if we were to point out that "being over the age of 108 is associated with a 92% increase in mortality".

References

European Court of Human Rights (2018). *Big Brother Watch and Others v. the United Kingdom - complaints about surveillance regimes: Press Release - Chamber Judgments*.

Gal, Tamas S. et al. (2014). "A data recipient centered de-identification method to retain statistical attributes". In: *Journal of Biomedical Informatics* 50. Special Issue on Informatics Methods in Medical Privacy, pp. 32–45. ISSN: 1532-0464.

Garfinkel, Simson L. (2015). "De-Identification of Personal Information". In: *NISTIR* 8053.

Hauskrecht, Milos et al. (2012). "Outlier Detection for Patient Monitoring and Alerting". In: *Journal of biomedical informatics* 46.1, pp. 47–55. ISSN: 1532-0464 1532-0480.

Marshall, Nathaniel S. et al. (2008). "Sleep Apnea as an Independent Risk Factor for All-Cause Mortality: The Busselton Health Study". In: *Sleep* 31.8, pp. 1079–1085.

Office for National Statistics (2017). "Crime in England and Wales: year ending June 2017". In: *Office for National Statistics Crime and Justice Statistical Bulletins*.

Oulasvirta, Antti et al. (2012). "Long-term effects of ubiquitous surveillance in the home". In: *UbiComp*.

Selinger, Evan (2014). "Philosopher Michael Lynch Says Privacy Violations Are An Affront To Human Dignity". In: *Forbes*.

Taboada-Crispi, Alberto et al. (2009). *Anomaly Detection in Medical Image Analysis*, pp. 426–446.

Thomas, Paul (2016). "Youth, terrorism and education: Britain's Prevent programme". In: *International Journal of Lifelong Education* 35.2, pp. 171–187.

Wachter, Sandra, Brent Mittelstadt, and Chris Russell (2017). "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR". In: *Harvard Journal of Law & Technology, 2018*.