Ethical advice report

Case
Type of SIM card as input variable for algorithmic
models that predict payment defaults

Source
Anonymized large multinational company with e-commerce platform

Summary advice
The audit commission advises against using type of SIM card as an
input variable in algorithmic models that predict payment defaults
and block afterpay services for specific customers. As it is likely
that type of SIM card acts as a proxy-variable for sensitive demo-
graphic categories, the model would run an intolerable risk of dis-
proportionally excluding vulnerable demographic groups from the
payment service. Absent reliable data that demonstrates otherwise,
the ethical risk of including the SIM card variable outweighs potential
benefits. The commission advises to consider a variety of alternatives
in dealing with payment defaults.

February 25th 2022

Algorithm
Audit

Problem statement - Predicting payment fraud
This document discusses ethical concerns regarding data-analyses and algorithmic methods used to predict payment fraud.

1. Introduction
For online purchases some companies offer customers a buy now and pay later service (afterpay). After having received the ordered products, some customers never pay the bill and don't respond to attempts by companies to settle a payment arrangement. Fraudulent customers are a considerable driver of costs for companies. Hence, companies use data-driven methods to detect potential fraudulent customers and to restrict the afterpay service for these costumers.

2. Data collection, analysis and algorithmic processing
Customers place orders online, either through an app on a mobile device or by using a web browser on a computer. Costumers leave traces on the company's platform which are collected in a dataset, consisting of profile characteristics, metadata and a label whether customers have defaulted the afterpay service. Profile characteristics comprise the customer's delivery address (per district), payment history and behavior on the company's platform. Metadata from mobile devices or web browsers are collected, e.g., type of SIM card and locations services. The data collection methods acquire highly reliable data. Missing datapoints occur occasionally and corresponding input data are excluded from the dataset. No further pre-processing interventions are applied to the collected dataset.

Data analysis methods are used on voluminous labeled datasets (typically >1 million observations) to gain insight in what type of customers are more likely to default the afterpay service. For example, correlation studies are conducted to shed light on potentially existing statistical relationships between input variables and the output variable. Additionally, supervised learning models are trained on historical data to classify new customers as risky or not risky. Such binary classification methods are operationalized to disable afterpay service for new customers identified as risky. That is, after having ordered products, this type of customer needs to complete the payment before the products are distributed to the delivery address.

3. Ethical concerns
Data analysis methods could indicate that the input variable 'type of SIM card' holds predictive power to identify potential fraudulent customers. However, it is known that the distribution of demographic groups in society differs across type of SIM-cards. For example, in The Netherlands Lebara and Lyca SIM-cards are relatively more often used by people with a Euro-African migration background, due to low intercontinental call charges. Investigative journalists of the Groene Amsterdammer revealed that the Dutch political party Denk uses Lebara and Lyca SIM cards as a filter for microtargeted Facebook ads. Companies are concerned about potential bias in classification algorithms that use the type of SIM card variable. Companies strive to develop

methods that use the potential predictive power of the SIM card variable while avoiding its ethical risks. As a standard baseline to examine algorithmic fairness, companies aim to test their binary classification methods (that allow or block the afterpay service) on so-called conditional demographic parity. This statistical measure is proposed by legal- and technical scientists to act as a minimal standard to examine the implicit bias of algorithms towards certain demographic groups. However, to perform this statistical measurement, data must be available to which demographic group customers belong, but exact data on this matter are unavailable. Estimation methods are considered to assign all customers a demographic group, for example based on location services. Though, these estimation methods face considerable challenges in terms of accuracy. Therefore, testing the companies' classification methods on conditional demographic density faces severe practical challenges. Hence, determining implicit bias in the used classification algorithm against certain demographic groups is not possible.

To put it short: Companies are worried that the type of SIM card variable could act as a proxy variable for demographic groups. If this is the case, prediction algorithms might develop an ethnic, religious or other demographic bias. The company's procedure on restricting afterpay service could then be perceived as discrimination. On the other hand, companies do not want to disregard relevant knowledge retrieved from historical data to deal with payment fraud.

Advice audit commission

Summary advice
The audit commission advises against using type of SIM card as an input variable in algorithmic models that predict payment defaults and block afterpay services for specific customers. As it is likely that type of SIM card acts as a proxy-variable for sensitive demographic categories, the model would run an intolerable risk of disproportionally excluding vulnerable demographic groups from the payment service. Absent reliable data that demonstrates otherwise, the ethical risk of including the SIM card variable outweighs potential benefits. The commission advises to consider a variety of alternatives in dealing with payment defaults.

Detailed advice

1. Disclaimer
First a disclaimer. The commission concludes that the case as described in the problem statement lacks relevant data and context to make an optimal, well-informed ethical assessment. In particular, statistical results about the model's performance, the default rate and a priori disparities across different types of SIM card would present a more complete picture of the case. Aside from statistical data, the commission would have benefited from relevant organizational context, such as the (business) reasons for developing the fraud prediction algorithm in the first place, possible alternatives and the consequences of not using an algorithmic model. Unfortunately, for this case it was impossible to obtain this data and information due to intrinsic and external limitations. The final advice must be perceived in light of this lack of information. The commission states that the availability of relevant data would enable the commission to better specify the ethical risks and the potential benefits of the predictive algorithm.

2. Ethical concerns
The commission advises that before deciding on the ethical risk of a specific predictor variable, an organization should have a clear rationale for deploying such a predictive algorithm at all. It is important to be aware of the fact that no model will be fully accurate and will always generate wrong predictions. Even before considering specific ethical risks of a predictive algorithm, such as the discriminatory effects, the developers of the algorithm should assess beforehand if the default rate and the financial loss for the organization is significantly high to outweigh the downsides of any algorithmic approach. Moreover, a prior assessment should demonstrate that the payment defaults are of such nature that they can only be prevented using a prediction model excluding certain users from the afterpayment service. That there is the technical possibility of using an algorithm to predict payment fraud is not a sufficient reason in itself. Algorithmic exclusion of people should not be used lightly. While the consequences of the exclusion are in the present case not as detrimental as, say, re-

fusing customers to buy products at all on the e-commerce platform, one should be careful not to underestimate the possible impact of such exclusions, especially when a discriminatory bias is suspected.

3. Proportionality of algorithmic method

The commission has pointed out that there might be other, less ethically precarious ways to deal with payment defaults. The model owners should then investigate root causes of payment defaults. For example, it might be the case that a concentrated, active criminal group is a main driver of payment fraud. There might be technical approaches to specifically target this criminal activity without applying a risk prediction algorithm on all users. Or it might be the case that a large fraction of payment defaults is not deliberately fraudulent, but the result of a confusing user interface, the lack of payment notifications and reminders, or the lack of barriers that make defaulting psychologically harder. In this case, changes to the interface would suffice. Another possibility would be to allow afterpayment service only to existing customers who have already made their first pre-payment. In the case at hand, it is not laid out sufficiently clear why the payment defaults warrant using a potentially biased prediction model and not an alternative, less risky approach. As a consequence, the potential benefits of using such a model do not outweigh the ethical risks associated with proxy-variables such as type of SIM card.

4. Proxy variables

The commission offers several reasons why using the variable type of SIM card is ethically precarious. It is dubious whether excluding certain people from a payment service based on type of SIM card (among other variables) is on firm legal ground when brought to court in a discrimination case. As this commission does not give legal but only ethical advice, we leave this aside. It is first of all likely that type of SIM card may act as a proxy variable for ethnicity and origin, as certain telecom providers target groups that make frequent international calls. To inspect this risk, it should be possible to give a rough estimate of the model's classification disparities among demographic groups, by using other proxy-variables, such as postal code and name. While these estimations remain inaccurate and should therefore never be used in the actual algorithm, for the developers they might serve as a warning sign whether demographic disparities occur in the model output. The case does not provide information or data on this. Hence, there is a risk that the prediction model is disproportionately excluding demographic groups through a disparate scoring of the various SIM cards. Moreover, we can assume that any differentiation based on SIM cards would at least partially be caused by the fact that is a proxy for underlying demographic variables. This is why including SIM card is ethically more precarious than, say, including income level, financial history or debt, which presumably does have a direct, intrinsic connection to payment default risk. A warning is in place regarding proxy variables that seem to contribute significantly to a model's performance. It may be hard to exclude such proxy variables when they seem indispensable to the algorithm. But the greater the contribution, the greater is generally the resulting possible bias.

5. Explainability and accountability of algorithmic model

In case organizations decide to indeed use a prediction algorithm to exclude certain users from afterpay-services, the commission advises to include a possibility of appeal. This implies, first, that excluded users should be informed about the fact that they have been excluded, preferably with the mention of the factors that contributed to the model's decision. And second, that the user should then be offered an accessible option of appeal. This means that a new decision needs to be made by a human person, who takes into consideration the user's reason of appeal and critically reviews the model's decision. This is not only a good ethical practice in itself, it also serves as a feedback loop for the organization to monitor the model's performance.

The above advice is the result of a deliberative group discussion. Audit commissions members should therefore not be held responsible individually for any of the above recommendations.

Audit commission facts

Date
The audit commission convened digitally on January 12th 2022. This report has been approved by all commission members and was affirmed on February 25th 2022.

Composition audit commission
- Fleur Jongepier, Assistant Professor Digital Ethics, Radboud University
- Jan Overgoor, PhD Management Science and Engineering, Stanford University
- Merel Noorman, Assistant Professor Law, Technology and Society, Tilburg University
- Roel Dobbe, Assistant Professor Engineering Systems and Services, TU Delft

Source of case
Anonymized large multinational company with e-commerce platform.