Case study – Fighting payment fraud

This document discusses ethical concerns regarding data-analyses and algorithmic methods used to detect payment fraud.

1. Introduction

For online purchases some companies offer customers a buy now and pay later service (after-pay). 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 after-pay 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 after-pay 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 after-pay 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 after-pay service for 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. 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 after-pay 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 after-pay service could then be perceived as discrimination. On the other hand, companies do not want to disregard relevant knowledge retrieved from historical data to fight payment fraud.

The question: To what extent and under what circumstances is it ethically justified to use the variable 'type of SIM-card' in data-analysis and prediction models to block after-pay services to specific costumers that are classified as risky? Are there ways to include the type of SIM-card variable in the data-analysis and prediction methods, while reducing the ethical risks?