Peer Review Framework for Predictive Analytics in Humanitarian Response

MODEL REPORT:
Cider model
Global Policy Lab at U.C. Berkeley

March 2022

OCHA CENTRE FOR HUMANITARIAN DATA

OCHA   centre for humdata

Model Report:
Cider model

1. Background

Cider is an open source toolkit developed by the [Global Policy Lab](#) at [U.C. Berkeley](#) for training models that use mobile phone data to infer poverty. Cider's main use case is predicting subscriber-level poverty from mobile phone metadata for the purpose of targeting social assistance or humanitarian aid. The output of the model is expected to be reliable for people who use their mobile phones at least occasionally (including people who use their phones fairly rarely).

Cider was initially implemented in Togo to [target humanitarian cash transfers](#) during the COVID-19 pandemic in November 2020 - August 2021. Cider will soon be released as an open source package for governments and NGOs to use similar targeting methods in future projects.

2. Main Findings and Recommendations

The main output of Cider is subscriber-level poverty predictions based on features derived from mobile phone data. Each output represents the inferred poverty of the individual subscriber, in the same units as the poverty measure on which the machine learning model is trained. Cider also automatically outputs information on the accuracy and fairness of the trained model.

The predictions generated by Cider are relevant to the time period in which phone data and training data labels are gathered. Therefore, if program implementers plan to use Cider to assign social protection benefits during an economic shock, Cider poverty scores from data collected pre-shock may not capture changes in poverty status resulting from the economic shock. In the Togo use case, training data and phone data for Cider were collected just prior to the cash transfer distribution (phone data was collected from March to October 2020, training data labels were collected in September and October 2020, and cash transfer dispersal was launched on November 1, 2020).

Privacy

Cider uses private and potentially sensitive information from individuals. Individuals should be given the option to opt in or opt out of mobile phone data usage. First, informed consent should be collected in surveys to merge survey responses to mobile phone metadata. Second, informed consent should be collected from individuals to process their mobile phone metadata upon program registration. Opt-in and opt-out workflows for collecting informed consent are included in the Cider toolkit. In addition to consent, data should be kept encrypted whenever possible. At a minimum, phone numbers included in mobile phone metadata should always be stored encrypted; we are also currently building features in Cider that will keep all mobile phone data encrypted until after informed consent is collected.

Transparency
The eligibility mechanism may not be transparent when phone data is used for poverty targeting. We are working on integrating Cider with LIME to provide individual explanations of predictions; it is also important that Cider users and policymakers consider how the eligibility mechanism is being made legible to program registrants.

You can find all the documentation regarding the model, its application and the review process at the following links:

- The Model Card describes version 1.0 of the model developed in November 2021.
- The Model Evaluation Matrix was completed in November 2021 by a technical expert from the Max Planck Institute for Human Development.
- The Implementation Plan was completed in December 2021. It summarizes the concrete actions that the model will trigger or inform.
- The Ethical Matrix aims to identify all stakeholders and potential issues regarding the intended use of the model. The Ethical Matrix was completed In February 2022 by an expert from the REACH Initiative.

A summary of the main findings and recommendations is provided below.

2.1 Technical Review

Intended use
- The quality and completeness of the mobile phone data being collected and shared can vary over time. We recommend updating the documentation with studies of effects of phone sharing availability.

Model development and documentation
- No specific recommendation provided other than updating the documentation with future case studies regarding model robustness when some features are missing.

Model Evaluation
- No action to be taken here.

Operational Readiness
- It is recommended to include more detailed guidance on privacy-sensitive handling, e.g. hashing, homomorphic encryption in future updates.

2.2 Ethical Review

False Negatives
Potential for exclusion because of false negatives. If the model produces false negatives, people who are poor enough to receive aid will not be eligible.

- We recommend reducing the impact of a false negative, if it occurs, by adding more information on system design exclusion error risk. For instance, more information about the survey methodology used or poverty measurement definition could be provided.

Systemic Bias
Potential for exclusion due to lack of phone access. Model requires access to phone in order for data for targeting model to be collected and for assistance to be provided.

- We recommend reducing the likelihood of systemic bias entering into model outputs by running face-to-face surveys to better understand the risk of systemic bias and the risk of excluding people without phones.

Lack of transparency
Difficulty for communities to understand targeting decisions, make complaints and hold their decision makers to account.

- We recommend increasing transparency of the model implementation and output by asking that informed consent from data subjects be included in the programme, prior to data collection and analysis.

Privacy
Potential risks coming from sensitive information contained in model's data.

- Cider uses private and potentially sensitive information from individuals. Individuals should be given the option to opt in or opt out of mobile phone data usage. First, informed consent should be collected in surveys to merge survey responses to mobile phone metadata. Second, informed consent should be collected from individuals to process their mobile phone metadata upon program registration.
- In addition to consent, data should be kept encrypted whenever possible. At a minimum, phone numbers included in mobile phone metadata should always be stored encrypted.

- Given the sensitivity of the data managed in the implementation of this model, we recommend the Client develop a data retention and destruction schedule that defines how long data should be retained and when and how it should be destroyed.