TOPIC:

Using Machine Learning and Mobile Phone Data in Program Targeting
Evidence from Anti-poverty Intervention in Afghanistan

Letizia Sormani
Anna De Seta

INTRODUCTION 3

THEORETICAL EXPLANATION 3

APPLICATION IN DEVELOPMENT ECONOMICS 4

Data and methods 5

LIMITATIONS 7

Limitations regarding the gradient-boosting method **7**

Limitation of the sample 7

Considerations on the combined model 8

CONCLUSION 8

BIBLIOGRAPHY 10

INTRODUCTION

In today's digitalized era, program targeting with machine learning and phone data has become a crucial force in reshaping how programs are structured and executed. First, it's important to highlight that targeting is a means of increasing program efficiency by increasing the benefit that the poor can get within a fixed program budget. Annually, hundreds of billions of dollars are spent on targeted social protection programs.

Machine learning is a subset of artificial intelligence that uses algorithms that continuously learn from and adapt to data. On the other hand, mobile phone data provides real-time information about the behaviors and interactions between consumers. Big data plays a crucial role in the diffusion and apprehension of this information.

The combination of these two technologies allows a more dynamic and innovative approach to determining eligibility, resource allocation, and decision-making.

The potential for application extends across various sectors, from public health initiatives to personalized strategies, this new approach with the use of these technologies can potentially improve how resources are allocated and how decisions are made by different companies. However, like any cutting-edge technologies and strategies, this presents some challenges along with opportunities. Limitations of these innovations will be highlighted and described later on in the report.

This paper focuses on how machine learning and mobile phone data can contribute to determining who should be eligible for program benefits. Evidence presented throughout the report, including insights from an anti-poverty intervention in Afghanistan, aims to present the practical implications of this innovative approach.

Apart from additional research, our main source is the paper written by Emily L. Aiken, Guadalupe Bedoya, Joshua E. Blumenstock, and Aidan Coville, titled "*Program Targeting with Machine Learning and Mobile Phone Data: Evidence from an anti-poverty intervention in Afghanistan*". The central theme of their research revolves around understanding the extent to which machine learning can precisely differentiate between ultra-poor households eligible for program benefits and those that are ineligible.

Their final conclusions will be that combining survey-based measures along with mobile phone data yields highly accurate classifications.

THEORETICAL EXPLANATION

Machine learning is defined as the "*use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data*".

Applying machine learning in the area of behavioral economics has introduced a new era of predicting and understanding human decision-making. Thus, it can assist companies in making better-informed decisions by trying to better interpret human choices based on individual preferences. Machine learning algorithms are crucial for behavioral economics to recognize patterns and non-linear relationships within vast data sets.

One of the most prevalent applications of machine learning is the analysis of preference and decision patterns. With the unveiling of behavioral trends, developing targeted marketing strategies, for example, becomes easier and more effective. Another crucial application of machine learning is personalized interventions based on the outputs generated by these algorithms.

The potential for innovation and influential applications is vast within the intersection between these two fields.

The foundation of machine learning is partly based on a model created by Donald Hebb in 1949, detailed in a book titled "The Organization of Behavior". Hebb's model can be described as a way of altering relationships based on the notion that when artificial neurons are activated together, their connection gets stronger compared to when they are activated separately. This concept helps artificial systems learn and adapt based on their experiences.

The study focuses on Afghanistan where call detail records (CDR) from a major phone operator are matched with data from the government's anti-poverty program.

The research connects two areas of study: one investigating the effectiveness of program targeting and the other exploring the possibilities of using non-traditional data with machine learning in low and middle-income countries, particularly in challenging environments such as Afghanistan.

APPLICATION IN DEVELOPMENT ECONOMICS

Delving into the practical applications of this methodology, we find a notable instance where machine learning was utilized to enhance program targeting efficiency in the study of the Targeting the Ultra-Poor (TUP) program implementation in Afghanistan.

The program we will analyze in depth took place between 2015 and 2018, targeting ultra-poor households in 6 specific Afghan regions. The initiative aimed to provide a time-limited large investment in productive assets (assets that can generate a cash flow, such as cows, sheep, or goats), access to savings accounts, temporary cash support, skill training, coaching, and other complementary services related to education and health.
These interventions addressed what were considered the baseline problems in rooted poverty, such as poor physical and mental health, food insecurity, financial constraint as well as low human capital.
In this case study, it is interesting to explore targeting as a tool rather than an objective, leveraging potential gains, errors, and costs. In fact, when designing a social assistance program, it is vital to consider the empirics that define eligibility in order to consciously evaluate the trade-offs. In particular, when implying new technologies, the goal is to improve the effectiveness of the program, identify target populations more accurately, allocate resources more efficiently, reduce administrative burdens, and ultimately enhance the delivery of services to those in need.

This research attempts to answer the extent to which non-traditional administrative data (such as mobile phone data), refined with machine learning techniques, can efficiently improve program targeting.

Data and methods

This objective is obtained by matching call detail records (CDR) with household survey data distributed by the government in targeting the ultra-poor program (TUP).
We start by defining the traditional methods employed to assess the eligibility for the social programs: proxy means test and community-based targeting.
Proxy means test is used to describe a situation where information on households correlated with welfare is used in a formal algorithm to serve as a proxy for household income.

Community-based targeting has at its base the core idea that the communities are given the responsibility for the allocation of externally assigned resources.

The eligibility for the program was assessed through a hybrid targeting method, hence a combination of proxy means test and community-based targeting, with traditional indicators, such as community wealth ranking (CWR), used to analyze the relative socioeconomic status in a given community, and a short follow-up in-person survey.
For a household to be considered ultra-poor, it had to be considered extra-poor in terms of CWR (ca. 43%) and to satisfy 6 different criteria:

1. financially dependent on women's work or begging
2. owing less than 800 meters square of land or living in a cave
3. primary woman under 50
4. no adult male income earners
5. school-age working children
6. no productive asset

Considering this classification method, 41% of more than 2500 households were eligible for the social program.

This paper considers the accuracy of three alternative methods, to then evaluate the best-performing program targeting approach by comparing results with the original recipients of the program. These alternative methods were built considering three different (traditional and non-traditional) measures of poverty:

1. asset-based wealth index, assessing the household's economic status through the ownership of various region-specific assets
2. consumption, assessing the household's economic status through the analysis of consumption patterns
3. CDR-based approach, assessing the household's economic status through inference by analyzing mobile phone usage.

In the latter approach, the new dataset, of 535 individuals, was built using the mobile data directly of the recipients of the TUP program. The collected findings contained detailed information on calls (duration and recipients), text messages (number and recipients), and recharges (time and amount).

For what concerns the handling of duplicates, in this case, households owning more than one phone, the approach was to consider the data matched to the head of the family. While considering missing values, in this case, households not owning any cell phones, the method used is random imputation, hence filling missing data with randomly assigned values.

In the context of machine learning, the handling of “special” data is important in data quality, hence impacting the model prediction, and bias reduction, avoiding incorrect conclusions.

Employing supervised machine learning, in this case, can help highlight a pattern in input features (in this case, phone trends) and output labels (categorizing the household as living above or below the poverty line)

In particular, the machine learning method employed is the gradient-boosting model. It is a supervised learning algorithm that attempts to accurately predict a target by combining the estimates of simpler (or weaker) learners, meaning relatively simple models, only slightly better at predicting than random guessing. The learners “learn” sequentially, and the final idea is to convert many weak learners to a single “complex” learner.

The architecture of gradient boosting is simple, starting with a naive model, we then make predictions. The model, then, employs the gradient descent procedure, which minimizes the loss (hence, residuals) when adding new learners to the ensemble, which is defined as the agglomeration of multiple prediction models. The data is then trained based on the newly built model, and the new model is added to the ensemble.

The accuracy was then evaluated through a ten-fold cross-validation technique, which divides the dataset into equal-sized subsets.

Results

To compare the performance of the different methods, it is useful to evaluate the confusion matrices: *Table 1* provides a detailed summary of the predictive performance.

In particular, we can read and evaluate the types of error (first type and second type) to assess model accuracy.

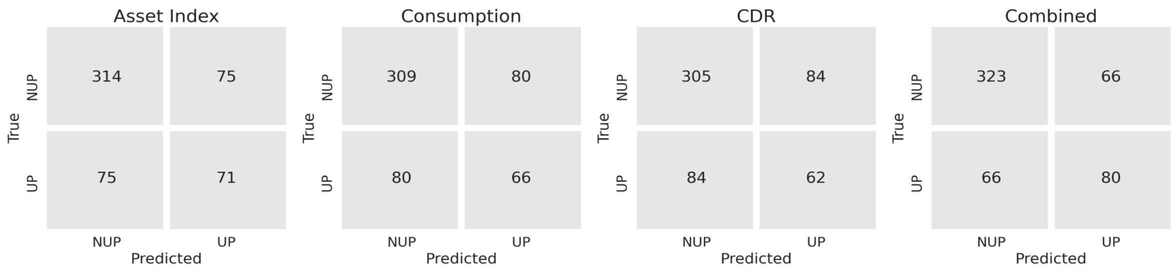

Table 1 - Confusion matrices

For the different indicators, we can additionally assess sensitivity by distinguishing false negatives, instances where the ultra-poor are incorrectly classified as non-ultra-poor, and true positives, instances where the ultra-poor are accurately classified as such.
Sensitivity is especially relevant when identifying positive cases is of utmost importance.
In terms of accuracy, the methodologies that utilize different indices yield comparable results. Notably, the combined targeting approach demonstrates superior performance, as evidenced by the lowest occurrence of both type I and type II errors.

LIMITATIONS

Considering this real-life example is useful to account for the possible limitations in the application of machine learning in program targeting.

Limitations regarding the gradient-boosting method

Considering the gradient-boosting machine learning method, the above-mentioned implementation naturally highlights some critical points. The application of this method will be sensitive to outliers and noise. Considering that each version of the model is built on previous residuals, outliers could shift disproportionally the attention on those points and cause overfitting.
Interpretability is another limitation, due to the high degree of complexity of the finalized model, as well as its non-linearity. This may not be considered a major drawback, especially in the specific use we are exploring with this paper, but quite relevant if transparency and fairness are expected from an external stakeholder.

Finally, machine learning largely depends on different parameters, and their tuning largely impacts the outcome and model efficiency. It is therefore important to conduct cross-sectional validation, which will additionally increase the computational cost of the model.

Limitation of the sample

In our specific example, in the application part, we have mentioned how missing values, as well as duplicates, were handled. This, of course, raises some questions regarding the validity of the approach. Households having multiple phones is an extremely indicative data of the possibility of not being ultra-poor. Considering only the head of the household is disregarding useful sample information. One possible approach could be to consider, for each duplicate, a fixed value for inputs. This would allow taking into account the additional information without creating the possibility of outliers.

The same could be said regarding the treatment of missing values, as random imputation decreases significantly the information power of the model.

Considerations on the combined model

One of the main conclusions of the study is the superior performance of the combined model. But a major drawback is the additional costs that this approach would require. The process of collecting personal data is costly, and it can defeat the purpose of using a non-traditional method such as machine learning, which indeed aims at enhancing convenience both in terms of cost and time. Additionally, the implementation of combined models raises some issues regarding privacy concerns and data collection. In fact, employing big data to assess eligibility bears important considerations, as the use of personal data encompasses cultural dimensions. It necessitates a thorough examination of legal aspects, particularly concerning data ownership and protection. Moreover, it underscores the importance of addressing sociocultural considerations explicitly in the context of these challenges.

CONCLUSION

In conclusion, the synergies between machine learning algorithms and mobile phone data contribute to an increase in efficiency and preciseness in determining program eligibility. This innovation has represented a groundbreaking shift in social protection initiatives, redefining the way in which these interventions are organized and executed.

The evidence drawn from the anti-poverty intervention in Afghanistan is tangible proof of the effectiveness of this new methodology. The accuracy with which ultra-poor households are classified as eligible is significantly enhanced.

However, while the application of machine learning to program targeting offers better insights into individual preferences and patterns, it is crucial to acknowledge its limitations and challenges. In particular, the costs of collecting consumption data may offset the potential cost-reduction benefits of machine learning.

Ongoing research and consideration guide the evolution of the transformative approach in the field of social protection programs.

BIBLIOGRAPHY

Aiken, E., Bedoya, G., Blumenstock, J., & Coville, A. (2022, June 22). *Program Targeting with Machine Learning and Mobile Phone Data: Evidence from an Anti-Poverty Intervention in Afghanistan*. ArXiv.org. https://doi.org/10.48550/arXiv.2206.11400

Grosh, M., Leite, P., Wai-Poi, M., & Tesliuc, E. (2022). Revisiting Targeting in Social Assistance: A New Look at Old Dilemmas. In *openknowledge.worldbank.org*. Washington, DC: World Bank. https://openknowledge.worldbank.org/entities/publication/a6b0063a-4805-5542-89a9-f6da877b5e37

Applying machine learning in behavioral economics. (n.d.). Www.linkedin.com. Retrieved February 6, 2024, from https://www.linkedin.com/pulse/applying-machine-learning-behavioral-economics-naveen-joshi/

Wealth Ranking | Sustainable Intensification Assessment Framework (SIIL). (n.d.). Sitoolkit.com. Retrieved February 6, 2024, from https://sitoolkit.com/the-five-domains/economic/poverty/wealth-ranking