Empowering the Underprivileged: How Machine Learning is Revolutionizing Humanitarian Assistance in Low-Income Countries

Abdullahi was her favorite grandson, Sontoy Mursal reflected. For all their life, they resided in a farming area in southern Somalia. In the past four years, rainfall has consistently fallen short of sustaining crop growth, forcing the small village to rely on outside goods. But it was different this year. The militant group, Al-Shabab, has taken over the region cutting off the movement of goods. Thus, Sontoy and her family must, every couple of months, embark on a 60-mile expedition by donkey cart to the city of Baidoa for food—a two-week journey.

Sontoy and her family are not alone. In the past year, 7 million people in Somalia—nearly half of the population—are deprived of urgent assistance for malnutrition, an issue well aware of the government. However, aid responses have been difficult and delayed. Al-Shabab has controlled southern and central Somalia, restricting the reach of humanitarian aid. Thus, the government lacked the ability to identify where assistance is needed the most. [3] They also cannot turn to government data to target the most impoverished as much of it is outdated. [2] In the case of Somalia, it has been almost 46 years since the last government census.

In fact, with limited data, Somalia could not officially declare a famine as it had no idea of the severity. [3] Without this declaration, global funding and assistance for the famine stalled as no one knew where to target their aid.[2] It took almost an entire year when finally a plan is in place to provide aid over the next rainy season as crops are expected to falter again. But for Sontoy and her family, it will be too late. Abdullahi died of starvation during one of the treks, forcing Sontoy to bury him on the side of a road—he was just six years old. [3]

Large-scale targeted humanitarian assistance programs, especially in times of crisis, have long been plagued by delays from the absence of immediate data. Even in the United States, a high-income country, the lack of available household data forced COVID-19 economic relief programs to rely on self-reported income leading to over $80 billion being given to fraudsters in 2021.[4] But for low-income countries, a disaster like this could lead to further permanent economic damage and—for families like Sontoy—lives. So, what's the solution?

It is evident that there is an essential need for up-to-date data to meet the demand for urgent humanitarian assistance. But repeatedly conducting censuses every year is extremely expensive and time-consuming for low-income countries. Additionally, for politically ravaged countries like Somalia, it's nearly impossible to conduct. Hence, the solution has turned to an alternative way to collect up-to-date statistics on these countries.

Mobile phone data usage has increased significantly in low-income countries over the past years. Mobile phone data or Call Detail Records (CDR) commonly can feature a user's call location, the call duration, the time of the call, and the user's data plans. With these features, CDR data can provide up-to-the-minute insight into a user's behavior. In fact, according to recent research at UC Berkeley, wealthy people use phones differently from the impoverished. Whom they call, where they text, which data plan they use, and when they call can all follow different patterns among different sociodemographic backgrounds.[2] Thus, researchers, like Dr. Aiken a professor at UC Berkeley, have turned to the incorporation of pattern recognition tools—like machine learning—to identify these patterns.

Machine learning (ML) algorithms can be trained from past data to recognize emerging trends and make inferences from them. A popular ML algorithm, the gradient-boosting model makes predictions or decisions on a set of input features. The model works by combining a large number of simpler models, called decision trees, in a way to improve its accuracy. A decision tree is like a flowchart that helps you make a prediction based on certain features. Take this as a simple example: imagine you are trying to predict the price of a house based on features like size, location, and number of bedrooms. If the house is located in a wealthy neighborhood, has more than 4 bedrooms, and is larger than 2000 square feet, the decision tree might predict a higher price given past data on houses with similar features. However, a single decision tree may not provide enough accuracy.

This is where gradient boosting comes in. Using many decision trees—each making its own prediction, gradient-boosting combines them until the model reaches a set desired accuracy. For instance, if multiple decision trees signify that wealthy neighborhoods always lead to higher prices, then the model will add more weight to that feature in predicting expensive housing. The model essentially learns from its mistakes and adjusts to improve its accuracy. A powerful and efficient technique that can offer a way to predict sociodemographic statistics.

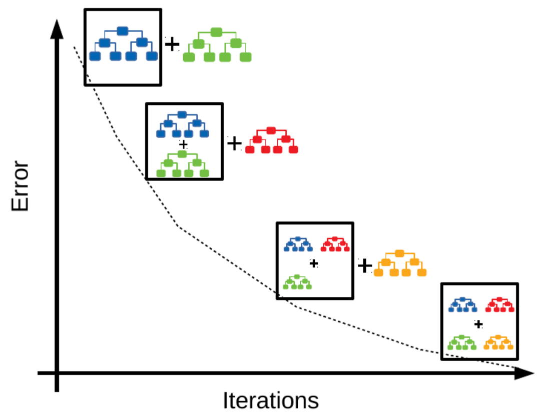

Figure 1: At each iteration, the gradient-boosting model combines a decision tree, improving its accuracy.

Dr. Aiken tested the viability of gradient-boosting in estimating poverty in Afghanistan with CDR. Like Somalia, Afghanistan's Targeting the Ultra-Poor (TUP) program has long been hindered by incomplete data on identifying poor households. Currently, the TUP program uses traditional estimation models from asset index and consumption data to recognize the ultra-poor. Due to the out-of-date nature of these sources, the process to collect data and train the models takes at least half a year. [1] However, CDR presents a readily accessible solution as it takes just hours to receive data. Furthermore, CDR allows the use of already gathered traditional data to supplement the model. Thus, along with a gradient-boosting model, CDR leads to a process that takes just a couple of weeks.

Poverty Estimation Models in Afghanistan

	Accuracy	Time Until Completion
CDR	69	~1-3 weeks
Asset Index	72	~6-9 months
Consumption	69	>1 year

Adapted from Aiken et. al (2020)

Figure 2: CDR-Based Model and Traditional Models Accuracy Rate[1] and Speed. *Created from DataWrapper*

The CDR-based gradient-boosting model successfully measured poverty in Afghanistan. The model, 69% of the time, could accurately differentiate a citizen who was impoverished from one who was not. The model was nearly as accurate as the best current TUP estimations method, however, took just weeks to build—a significant result.[1] Thus, CDR and a gradient-boosting model provide an approach to targeting humanitarian aid in low-income countries that can meet the urgency of a crisis. But what does it look like in action?

With the peak of the COVID-19 pandemic in 2020, Togo's government turned to this machine learning approach in implementing its cash aid program. Dr. Blumenstock from the School of Information at UC Berkeley partnered with Dr. Aiken in adapting a gradient-boosting model for Togo's economic condition. Similar to Somalia, Togo's last conducted census was over a decade ago. Thus, the researchers conducted a two-week-long phone survey in September 2020 to collect information on the living conditions of 15,000 households. Then, the researchers trained the gradient-boosting model, along with existing CDR, to recognize characteristics of people living on less than $1.25 per day—the threshold for cash aid. Togo's model accurately predicted the target threshold 70% of the time. Subsequently, in November, less than two months after the start of the research, the cash assistance program was rolled out. The method was significantly faster and more accurate than any other modeling methods considered by Togo's government. Currently, the program has provided nearly $10 million to around 137,000 of the country's poorest citizens. [2]

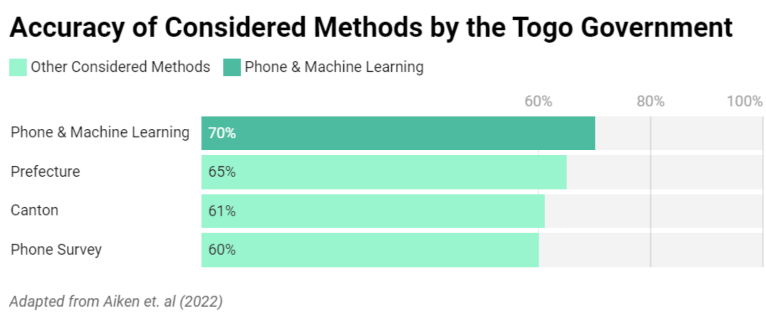

Figure 3: Accuracy of Considered Methods by the Togo Government[2]
Created from DataWrapper

The use of mobile phone data and machine learning in providing estimates for targeted humanitarian aid programs in low-income countries has the potential to revolutionize aid delivery. Therefore, as the trends of machine learning and phones continue to grow, there must be a rising awareness of the potential impact it can make on empowering impoverished communities. By leveraging mobile phone data and machine learning, governments and aid organizations can better understand the needs of the populations at a much more rapid rate. All of this allows aid programs to prepare and respond more efficiently—potentially saving many lives and preventing tragedies like Sontoy's family.

References

[1] Aiken, E. L., Bedoya, G., Blumenstock, J. E., & Coville, A. (2020). Program targeting with machine learning and mobile phone data: Evidence from an anti-poverty intervention in Afghanistan. *Journal of Development Economics*, *161*, 103016. https://doi.org/10.1016/j.jdeveco.2022.103016

[2] Aiken, E., Bellue, S., Karlan, D., Udry, C., & Blumenstock, J. E. (2022). Machine learning and phone data can improve targeting of humanitarian aid. *Nature*, *603*(7903), 864–870. https://doi.org/10.1038/s41586-022-04484-9

[3] Jerving, S. (2022, September 6). *'The Cavalry Hasn't Arrived:' Somalia on the Brink of Famine*. 'The cavalry hasn't arrived:' Somalia on the brink of famine. Retrieved March 15, 2023, from https://devex.shorthandstories.com/the-cavalry-hasn-t-arrived-somalia-on-the-brink-of-famine-brink-of-famine/index.html

[4] Office of Comptroller General of the United States. (2023). *Emergency Relief Funds* (GAO Publication No. 23-106556). U.S. Government Accountability Office. https://www.gao.gov/assets/gao-23-106556.pdf