Teaching Old Data To Do New Tricks:
Machine Learning with Facility and Population Data in Uganda

Amy Finnegan, PhD, David D. Potenziani, PhD, Cyrus Elahi, MSc, Nicholas Matsiko, Irene Wanyana, and Caroline Karutu, PhD

IntraHealth INTERNATIONAL
Because Health Workers Save Lives.

6340 Quadrangle Drive, Suite 200
Chapel Hill, NC 27517
1 (919) 313-9194
afinnegan@intrahealth.org

Goal

Understand machine learning (ML) data needs and possibilities in the context of improving the quality of integrated service delivery in Ugandan health facilities.

Week 1

We started by gaining buy-in from our USAID-funded Regional Health Integration to Enhance Services in Eastern Uganda (RHITES-E) team and the Ministry of Health. Then we assembled an interdisciplinary team including:

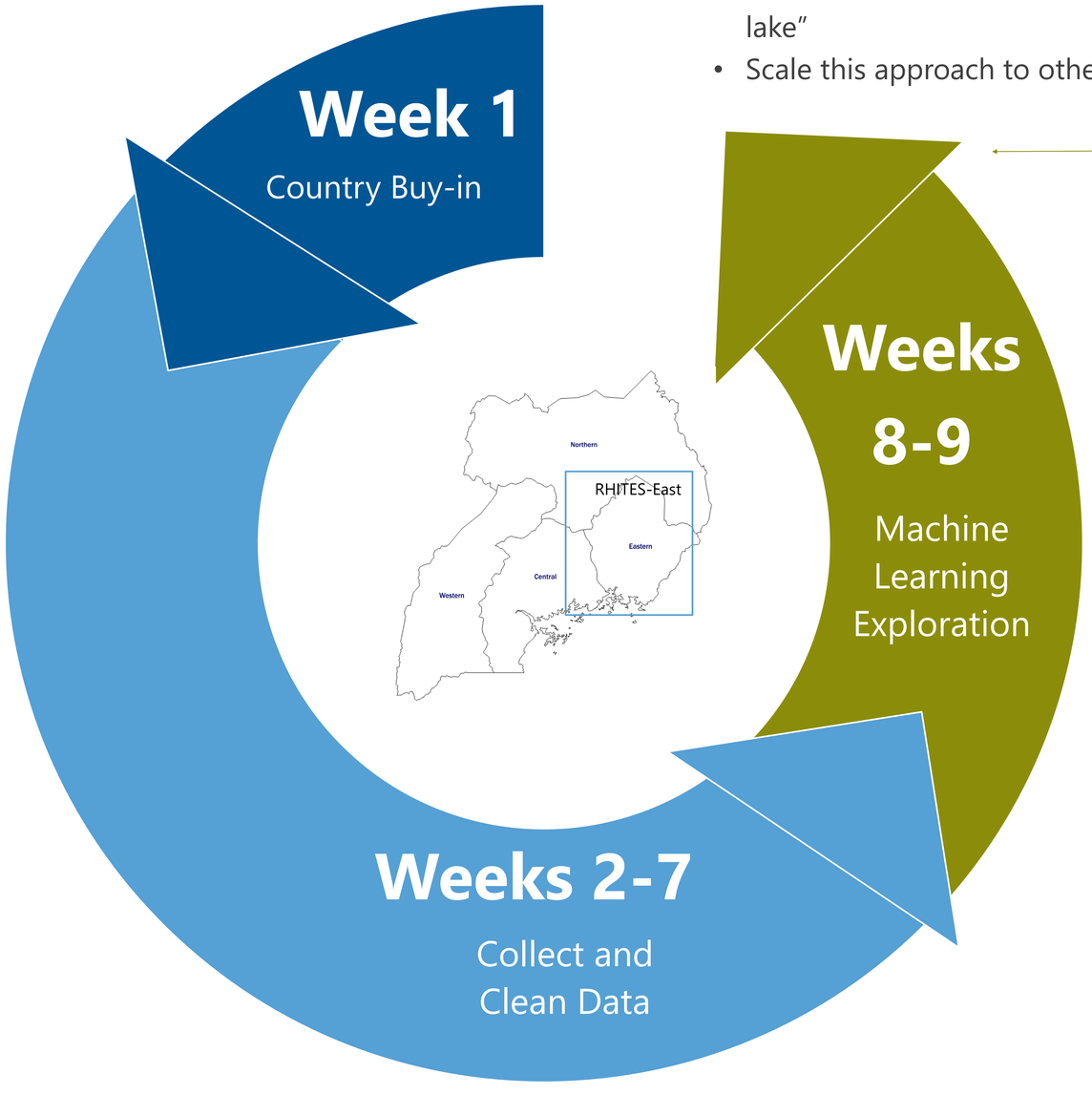
- IntraHealth Program Manager
- Machine Learning Technical Lead
- Two master's level students with strong ML backgrounds (Uganda and US-based)
- RHITES-E M&E Officer based in Uganda.

Weeks 2-7

We needed to create a "data lake" for analysis, so we collected 1,000s of variables from:

- IntraHealth's RHITES-E project
- Uganda's District Health Information System (DHIS2); Human Resources Information System (iHRIS); Demographic & Health Survey (DHS); and census.

Next, we had to clean and pre-process the data to prepare them for analysis. We learned that not all data were useful. We cut the number of variables included in the model from 1,200 to 472.



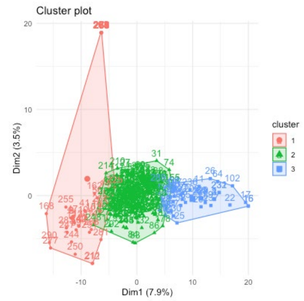
Conclusion

We learned that machine learning is a journey, not a destination. This activity was just the first step—an initial data collection, cleaning, and analysis.

In the next several months we plan to:

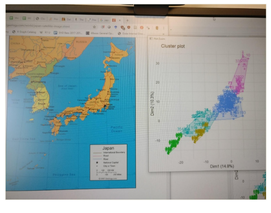
- Work with the RHITES-E team to generate hypotheses to test in the "data lake"
- Scale this approach to other countries we work with.

Weeks 8-9



We identified three clusters of facilities in the data.

Our final model included **38** variables that correctly predicted cluster membership. Variables included markers of facility quality.



We tried a few early models and ended up visualizing Japan. Oops! We learned the ML process is very iterative.

Acknowledgements: *We are grateful for the support provided by our Ugandan colleagues, the RHITES-E project, and the digital health team at IntraHealth International including Wayan Vota and Dana Acciavatti.*