THE UNIVERSITY OF CHICAGO
DATA SCIENCE INSTITUTE

The International Rescue Committee (IRC) is a non-profit organization focused on providing humanitarian aid to individuals in regions affected by conflict. As a part of the organization's Mali childhood malnutrition treatment program, the IRC asked the University of Chicago's Data Science Clinic to develop a model to predict whether a patient will default from treatment. Using the model, the IRC can deliver targeted support to patients at risk of defaulting, increasing the likelihood of treatment completion.

The team faced challenges with confirming patient villages and handling an unbalanced number of defaulters and non-defaulters. The dataset included patients with unrealistic travel distances between their villages and health facilities, as well as an inconsistent yearly default rate. The team collaborated with the IRC to handle data irregularities and applied data balancing techniques to address data imbalance.

Predictive factors such as specific health centers were identified through the efforts of the team. After testing various methods, the best model developed was a tuned algorithm optimized to handle the data's many categorical features (CatBoost). This final model, trained on patient admission data, correctly identified 42% of defaulters.

Figure 1: Confusion Matrix of Recommended Model