Predicting Poverty

The goal is to demonstrate your knowledge of machine learning by predicting whether an individual household is poor. There are two data sets provided, train and test, where the train data set is the training data to build the model and the test data will be used for scoring. You will notice that the column names and factors are encoded, therefore it is important to rely on machine learning concepts rather than ideas based in economic principles to improve your model. No outside information should be used to train the model. The scoring metric used will be Logloss according to

$$\frac{1}{N-1}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

where

$N : the\ total\ number\ of\ observations$
$y_i : the\ actual\ result\ for\ observation\ i$
$\hat{y}_i : the\ probability\ of\ the\ household\ being\ poor$