Household Poverty Level Prediction

TEJASWINI TALEKAR | KEVAL VANKUDRE

Problem Statement:

Many social programs have a hard time making sure the right people are given enough aid. It's especially tricky when a program focuses on the poorest segment of the population. The world's poorest typically can't provide the necessary income and expense records to prove that they qualify. Thus, we need consider numerous tangible factors for the given Costa-Rican population.

Objective:

The objective is to predict poverty on a **household level**. We are given data on the individual level with each individual having unique features but also information about their household. In order to create a dataset for the task, we'll have to perform some *aggregations of the individual data* for each household. Moreover, we have to make a prediction for every individual in the test set, but *"ONLY the heads of household are used in scoring"* which means we want to predict poverty on a household basis.

Dataset Details:

The file is a set of household characteristics from a representative sample of Costa Rican Households. The dataset has observations for each member of the household but the classification is done at the household level.

Number of rows: **9557**

Number of Columns: **143**

Steps taken:

The end objective is a machine learning model that can predict the poverty level of a household. However, before we get carried away with modeling, it's important to understand the problem and data. Also, we want to evaluate numerous models before choosing one as the "best" and after building a model, we want to investigate the predictions. Our roadmap is therefore as follows:

- Understand the problem (we're almost there already)
- Exploratory Data Analysis
- Feature engineering to create a dataset for machine learning
  - Backward Elimination
  - Principal component analysis
- Compare several baseline machine learning algorithms. Algorithms used:
  - K - Nearest Neighbors Classifier
  - Artificial Neural Network
  - Random Forest Classifier
- Optimize the selected model
- Investigate model predictions in context of problem
- Draw conclusions and lay out next steps

References:

- https://www.kaggle.com/ashishpatel26/costa-rican-household-poverty-level-prediction
- https://www.kaggle.com/c/costa-rican-household-poverty-prediction/data