

# Identify the level of income qualification

By:Krishna Aryal

Trainer:Abhishek



# Data understanding

Variables:

Income level of families of Latin America was calculated.

1. Train data has 143 columns and test data has 142 columns.
2. Target is output variable
3. There are 130 columns integer types, 8 float64 type, that won't create problem.  
Let's figure out object type column and what are they.

# Model used

1. Predict the accuracy using random forest classifier.
2. Check the accuracy using random forest with cross validation.

# Model Interpretation and findings

- 1 = extreme poverty 2 = moderate poverty 3 = vulnerable households 4 = non vulnerable households very small value of small poverty in comparison with others shows there is biasness in the data.
- 85 families do not have same poverty level
- Out of 9557 rows 2973 are head of families. idhogar, Household level identifier helps to figure out families without head.
- Target is calculated using Random forest classifier.
- Model was later validated by cross validation with 10 random forest.
- Mean of prediction score is later calculated.