

Income Qualification

Income Qualification is a project that uses the data provided in the dataset regarding the Latin American Families and their household , education level with other criteria that might reflect on the income of individuals of the families.

Problem Statement Scenario:

Many social programs have a hard time making sure the right people are given enough aid.

It's tricky when a program focuses on the poorest segment of the population. This segment of population can't provide the necessary income and expense records to prove that they qualify.

In Latin America, a popular method called Proxy Means Test (PMT) uses an algorithm to verify income qualification. With PMT, agencies use a model that considers a family's observable household attributes like the material of their walls and ceiling or the assets found in their homes to classify them and predict their level of need.

While this is an improvement, accuracy remains a problem as the region's population grows and poverty declines.

The Inter-American Development Bank (IDB) believes that new methods beyond traditional econometrics, based on a dataset of Costa Rican household characteristics, might help improve PMT's performance.

Following actions should be performed:

- Identify the output variable.
- Understand the type of data.
- Check if there are any biasness in your dataset.
- Check whether all members of the house have the same poverty level.
- Check if there is a house without a family head.
- Set poverty level of the members and the head of the house same in a family.
- Count how many null values are existing in columns.
- Remove null value rows of the target variable.
- Predict the accuracy using random forest classifier.
- Check the accuracy using random forest with cross validation.

After we perform the above mentioned operations . We clean up the data and apply the Random Forest Classifier which then predicts and helps us to plot the prediction onto a graph , which gives us an Idea about the amount of every feature we have in our data set needed for Income qualification .

The graph looks as follows:

