

## Problema predictivo

Dadas las características de la vivienda (materiales con el que esta hecha la casa, el estado de las paredes, estado del techo, número de habitaciones, etc.), si la casa cuenta con servicios (si tienen energía eléctrica, alcantarillado, suministro de agua, etc.) y características de la persona que habitan en la casa (genero, estado civil, edad, cantidad de personas que habitan en la casa, escolaridad, etc.), el objetivo es predecir a que grupo de niveles de ingresos (1: pobreza extrema, 2: pobreza moderada, 3: hogares vulnerables, 4: hogares no vulnerables) pertenece un registro.

## Dataset

- **Nombre:** Costa Rican Household Poverty Level Prediction
- **Tipo:** Competición
- **Link:** <https://www.kaggle.com/competitions/costa-rican-household-poverty-prediction/overview>
- **Registros:** 9557
- **Columnas:** 143

## Métricas de desempeño

- **Machine learning:** Como métrica de machine learning utilizaremos el F1-score definido por la competencia.

Supongamos que se trabaja para una organización sin animo de lucro que brinda asistencia económica a hogares en situación de pobreza a la población de costa rica. El objetivo de la organización es asignar recursos de manera eficiente para ayudar a las familias que más lo necesitan. Esta organización cuenta con recursos limitados.

- **Negocio:** Mirando el problema desde una perspectiva de negocio se busca que a partir de la caracterización de una vivienda y los miembros que habitan en esta, para ello se debe tener en cuenta lo siguiente:
  - Es crucial garantizar que ningún hogar en situación de pobreza extrema sea pasado por alto. El costo de cometer un falso negativo puede ser muy alto, ya que significa que una familia necesitada podría no recibir la ayuda necesaria en el momento adecuado.
  - Es esencial asignar los recursos de manera efectiva, ya que, de lo contrario, estas ayudas podrían llegar a hogares que no se encuentran en situación de pobreza extrema. Esto no solo resultaría en un desperdicio de recursos, sino

que también podría reducir la capacidad de ayudar a aquellos que realmente lo necesitan en el futuro.

## Desempeño deseable en producción

- **Recall:** La organización sigue de cerca el recall para garantizar que se estén identificando la gran mayoría de los hogares en pobreza extrema. Un recall del 90% significa que se identifican correctamente el 90% de los hogares que realmente están en esta situación.
- **Precision:** La organización supervisara la precision para asegurarse de que la mayoría de los hogares identificados como en pobreza extrema realmente lo sean. Una precision del 90% significa que el 90% de los hogares identificados como en extrema pobreza lo son realmente.
- **F1-Score:** El F1-Score se utiliza como una métrica general para evaluar el equilibrio entre recall y precision. Se busca alcanzar un F1-Score de al menos 90% para asegurar un equilibrio entre identificar hogares en necesidad y asignar recursos con precisión.

Un estudio de mercado exhaustivo en Costa Rica reveló que aproximadamente el 60% de la población del país se encuentra en situación de pobreza. Esta cifra alarmante subraya la urgente necesidad de implementar medidas efectivas para abordar esta problemática social. Dada la alta prevalencia de la pobreza en Costa Rica, es esencial maximizar la efectividad de las iniciativas de ayuda. Un objetivo del 90% en Recall, Precisión y F1-Score asegura que el modelo identifique con precisión y brinde apoyo a una proporción significativa de los hogares en necesidad, contribuyendo directamente a la reducción de la pobreza en el país.