REPORT ON POVERTY PREDICTION USING PSLM-2020 DATASET

Jawad Ahmad

What is the PSLM Survey?

The Pakistan Social and Living Standards Measurement (PSLM) Survey is a national survey that collects information about the lives of people in Pakistan. The PSLM-2020 dataset focuses on important topics like education, health, jobs, income, and how people live. It includes data from both cities and villages, covering thousands of families to represent the whole country. This dataset helps understand how much people earn, what jobs they do, and how many are living in poverty. It's very useful for creating plans and policies to improve the lives of people in Pakistan

Why Understanding Poverty Matters?

Poverty remains a critical global challenge, and understanding its determinants is essential for effective policy-making. The World Bank defines poverty as a state of deprivation where individuals lack the resources to meet their basic needs. Using this definition as a guiding principle, the PSLM-2020 dataset was explored to predict and estimate poverty levels. This report outlines the steps taken to process the dataset, prepare it for analysis, and implement predictive modeling.

DATA EXPLORATION AND CLEANING

Columns in Section E (Employment and Income)

The dataset's structure was examined, particularly focusing on Section E, which pertains to employment and income data. Columns related to income and employment (seaq07 to seaq15) were identified for further analysis.

Handling Missing Values

Missing values in income-related columns were replaced with 0 to simplify calculations and ensure that all rows had complete income data for analysis.

FEATURE ENGINEERING

Annual Income Calculation

Main Occupation Income: The code calculates annual income from the main occupation using either the reported monthly income multiplied by months worked or the directly provided annual income, depending on the reporting format (seaq07, seaq08).

$$\text{Main Annual Income} = \begin{cases} \text{Monthly Income} \times \text{Months Worked}, & \text{if Report Period} = 1 \\ \text{Reported Annual Income}, & \text{otherwise} \end{cases}$$

Secondary Occupation Income: A similar approach is applied to the secondary occupation income (seaq11, seaq12).

$$\text{Secondary Annual Income} = \begin{cases} \text{Secondary Monthly Income} \times \text{Months Worked,} & \text{if Secondary Report Period} = 1 \\ \text{Reported Secondary Annual Income,} & \text{otherwise} \end{cases}$$

Total Annual Income

Total annual income is calculated by summing the main and secondary occupation incomes for each household. This composite metric provides a holistic view of household earnings.

$$\text{Total Annual Income} = \text{Main Annual Income} + \text{Secondary Annual Income}$$

POVERTY CLASSIFICATION

World Bank Threshold: The poverty threshold is set at \$2.15/day, translating to \$784.75/year. This threshold is applied to categorize households as either "poor" or "non-poor."

Binary Classification: Households with a total annual income below the threshold are labeled as "poor," and those above it as "non-poor."

For a household earning \$1500/year:

$$1500 > 784.75 \implies \text{Non-Poor} 1500 > 784.75$$

For a household earning \$700/year:

$$700 < 784.75 \implies \text{Poor} 700 < 784.75$$

DATA VISUALIZATION

A bar chart is generated to visualize the distribution of poverty status across households. The chart highlights the number of households classified as "poor" versus "non-poor," providing an intuitive understanding of the dataset's poverty profile.