Assignment-4

Analysis Objective.

The objective of this challenge is to create a machine learning model to predict whether an individual earns above or below a certain amount.

This solution can potentially reduce the cost and improve the accuracy of monitoring key population indicators such as income level in between census years. This information will help policymakers to better manage and avoid income inequality globally.

Data background.

There are ~200 000 individuals in train and ~100 000 individuals in the test file.

The train & test data will be used to create a machine learning model to predict if an individual earns above 50 000 of a specific currency.

The key variables are as follows:

- Age.
- Gender.
- Education.
- Class.
- Education institute.
- Marital status.
- Race.
- Is hispanic.
- Employment commitment.
- Unemployment reason.
- Employment state.
- Wage per hour.
- Is part of labor union.
- Working week per year.
- Main Industry code.
- Main Occupation code.
- Total employed.

- Household summary.
- Under 18 family.
- Veteran benefit.
- Tax status.
- Gains.
- Losses.
- Stocks status.
- Citizenship.
- Residence 1 year ago.
- Old residence state.
- Income above limit.

Based on the variables set up and the data target requirements, the analysis can be assumed to be based on 20th century American population data where the median income was about $ 50,000.

Why Predict Income?

Income prediction extracts insights from individual and population-level data as it offers the ability to forecast income levels, assess financial risks, target marketing campaigns, and inform crucial decision-making in diverse spheres. However, ethical considerations, potential biases, and data privacy concerns demand careful attention alongside its undeniable benefits.

Importance in Public Policy:

- Predicting tax revenue and allocating government resources based on income distribution.
- Identifying individuals in need of social assistance programs based on income poverty.
- Evaluating the effectiveness of government policies aimed at income inequality.
- Designing progressive taxation systems based on predicted income levels.