**Problem Statement & ML Flow**

**Understand the data and variables:**

The adult dataset contains demographic and income information about individuals.

**Feature columns:**

1**. age**: The age of the individual.

2.**workclass**: The employment sector of the individual.

3.**fnlwgt**: Final weight,while doing servey.

4.**education**: Highest level of education acheived.

5.**education-num** : Numerical version of education.

6.**marital-status**: Marriage status.

7.**occupation**: Type of occupation.

8.**relationship**: Relationship status.

9.**race**: The race of the individual.

10.**sex**: The gender of the individual.

11.**capital-gain**: Income from capital gains.

12.**capital-loss**:Income from capital loss.

13.**hours-per-week**: Number of hours worked per week.

14.**native-country**: The country of origin.

**Target columns:** income column contains values like <=50k and >50k.

**CategoricalVariables:**workclass,education,marital-status,occupation,relationship,race,

sex, and native-country.

**Numerical Variables:**age,fnlwgt,education-num,capital-gain,capital-loss,hours-per-week.

**Target Variables:**income.

**Define Problem Statement:**

The task is to build a predictive ML model that determines whether an employee earns more than 50k annually by analysing the given feature columns like demographic columns, job type, education level, and other columns.

**Machine Learning Flow:**

1.Form a Hypothesis

2.Collect and Explore the data

3.Data cleaning and Preprocessing

4.Data Splitting

5.Model Selection

6.Model Training

7.Model Evaluation

8.Hyperparameter Tuning

9.Model Deployment

10.Monitoring and Maintenance

About Dataset:

This is UCI Machine Learning Repository’s Adult dataset. There are 4 files in the folder., adult.data : This is the main dataset file for ML model . This incudes demographic information with income variables for prediction.

Adult.names : This helps to understand the structure of dataset, includes attributes (feature names) and variables.

Adult.test : This is testing data used for validating the ML model.

Index :

Old.adult.names : This might be older version of structured dataset.

Objective of ML model : To predict the people who make over $50k per year based on all factors given.