Income Prediction

Under Mentorness Internship Program

Problem Statement:

The objective of this machine learning task is to construct a predictive model for classifying individuals based on whether their income exceeds a specified threshold of $50,000. The provided dataset encompasses a variety of demographic, socio-economic, and employment-related features for a diverse group of individuals. The goal is to leverage this information to create a classification model that effectively discerns between individuals with incomes above and below the $50,000 limit.

Dataset Description:

1. Demographic Information:

- Age: Age of the individual.
- Gender: Gender of the individual.
- Education: Educational background of the individual.
- Marital Status: Marital status of the individual.
- Race: Ethnic background of the individual.
- Hispanic Origin: Whether the individual is of Hispanic origin.

2. Employment Details:

- Class of Worker: The class of worker the individual belongs to.
- Enrolment in Educational Institution: Whether the individual is enrolled in an educational institution.
- Employment Commitment: Full or part-time employment status.
- Unemployment Reason: Reason for unemployment.
- Employment Status: Whether the individual owns a business or is self-employed.
- Wage per Hour: Hourly wage of the individual.
- Labour Union Membership: Membership status in a labour union.
- Weeks Worked in a Year: Number of weeks worked in a year.
- Industry and Occupation Codes: Codes indicating the industry and occupation of the individual.

3. Financial Information:

- Gains: Financial gains.

- Losses: Financial losses.
- Dividends from Stocks: Dividends received from stocks.

4. Migration and Residence Details:

- Citizenship: Citizenship status of the individual.
- Migration History: Information on migration.
- Residence Changes: Changes in residence.

5. Family and Veteran Status:

- Household and Family Statistics: Detailed household and family statistics.
- Veterans Administration Questionnaire: Responses to the questionnaire for Veterans Administration.
- Veteran Benefits: Benefits received by veterans.

Target Variable:

- Income Above $50,000: Binary classification indicating whether the individual's income is above $50,000.

Challenges:

- Handling missing data and outliers.
- Dealing with categorical variables and encoding them appropriately.
- Identifying relevant features that contribute to accurate income predictions.
- Balancing the dataset to avoid bias towards the majority class.

Evaluation Criteria:

- The model will be evaluated based on metrics such as accuracy, precision, recall, and F1 score.
- Emphasis on the ability to correctly predict individuals with income above $50,000.

Income Prediction

Under Mentorness Internship Program

Deliverables:

- Trained machine learning model.

- Evaluation metrics and analysis.

- Documentation on feature importance and model insights.

Expected Outcome:

- A robust and interpretable model capable of predicting income levels, providing valuable insights for decision-making in various domains such as finance, social services, and policy planning.