### Abstract

The "GP\_PROJECT\_Better performance" project primarily focuses on applying Genetic Programming (GP) techniques for predictive modeling. The project encompasses various stages of data processing, including cleaning, preprocessing, and feature encoding, before implementing a GP-based model. The use of custom functions and tailored GP parameters highlights an innovative approach to handling specific data types and problem-solving strategies. The project culminates in the evaluation of the GP model's predictive performance and the visualization of the results.

### Introduction

Genetic Programming, a branch of evolutionary computation, provides a framework for evolving programs and models to solve specific problems. This project leverages GP to develop a predictive model, likely aimed at income prediction or a similar classification task.

### Methodology

1. **Data Loading and Cleaning**: The project starts with loading two datasets and performing initial data cleaning, which involves handling missing values and ensuring data quality.
2. **Data Preprocessing**: This step includes encoding the target variable and one-hot encoding of categorical features, preparing the data for effective modeling.
3. **Genetic Programming Setup**: The core of the project is setting up the GP environment. This involves defining custom functions for data handling, setting GP parameters (population size, crossover and mutation probabilities, etc.), and initializing the GP primitives and fitness function.
4. **Model Development and Evaluation**: The GP model is trained on the preprocessed data. The fitness function, based on accuracy score, guides the evolution of the GP individuals. Predictive models are then evaluated on their performance.

### Results and Discussion

The project successfully implements a GP-based model, evaluating its performance through accuracy metrics. Predictions are made on test data, and results are visualized using histograms to understand the distribution of predicted classes. This visualization aids in interpreting the model's effectiveness and potential biases.

### Conclusion

The "GP\_PROJECT\_Better performance" project demonstrates a novel application of Genetic Programming in predictive modeling. The project successfully navigates through data preprocessing, GP model development, and result analysis, showcasing the potential of GP in machine learning tasks. However, the project could benefit from additional documentation and explanatory text within the notebook to enhance its comprehensibility and utility for future users or researchers.

### Future Work

Further work could explore the optimization of GP parameters, the inclusion of more complex fitness functions, and the application of the model to other datasets or prediction tasks. Additionally, incorporating detailed documentation and analysis would significantly enhance the project's accessibility and educational value.