Predicting Inequality Peaks: Machine Learning Projections in the World of Wealth Dynamics

Background:

Income inequality is a critical socio-economic indicator that affects nations' overall well-being. The Gini Index is widely used to measure income distribution within a country, with a higher Gini Index indicating greater inequality. In this hackathon, the goal is to develop a machine learning model that can accurately predict the Gini Index based on various socio-economic factors.

Dataset:

The dataset provided contains information on various socio-economic indicators for different countries over multiple years. These indicators include population, GDP, regional information, income group, and a range of percentile values (p1 to p100) representing income distribution.

Task:

Build a web app to predict the Gini Index for a given country and year using the provided dataset.

Evaluation:

The model's performance will be evaluated using appropriate regression metrics, such as Mean Squared Error (MSE) and R-squared. The goal is to create a model that generalizes well to unseen data and accurately predicts the Gini Index.

Data Columns:

- Independent Variables: Various socio-economic indicators (e.g., population, GDP, regional information, income group, percentiles p1 to p100).

- Target Variable: Gini Index (column name: 'gini').

Deliverables:

Participants are expected to submit:

1. Model Building Notebook: A well-structured notebook detailing the step-by-step process of building the machine learning model for predicting the Gini Index. Include clear explanations, code documentation, and any relevant visualizations.
2. Dataset: The dataset used for training and evaluating the machine learning model. Ensure that the dataset is properly documented, and any preprocessing steps are clearly outlined.
3. Web App: A fully functional web application developed using Flask for model deployment. The web app should provide an intuitive interface for users to interact with the Gini Index prediction model.
4. PPTS (Presentation Slides): A comprehensive set of presentation slides outlining the key aspects of the project. Include sections on data analysis, model architecture, evaluation metrics, insights gained, and the deployment process. The presentation should be clear, visually appealing, and suitable for sharing during the project demonstration.Conduct thorough data analysis to comprehensively understand the dataset, extracting meaningful insights that inform the development and refinement of the machine learning model for Gini Index prediction.

Additional Information:

- Participants are encouraged to explore feature engineering and model hyperparameter tuning to improve model performance.

- Consideration should be given to potential biases and ethical considerations in the prediction of income inequality.

Dataset Source:

The dataset for this hackathon:
https://docs.google.com/spreadsheets/d/10G_CVXG9qV3GztIFQ0WDGniIP4uIaoMT/edit?usp=drive_link&ouid=101673839528144597155&rtpof=true&sd=true

Data Description: https://www.wider.unu.edu/sites/default/files/WIID/PDF/WIID-User-Guide-30JUN2022.pdf