Assignment-23.02.2026

Experiment: Income Pattern Data Analysis Using Visualization and Linear Regression

Problem Statement:

You are provided with an 'Income Pattern Dataset' containing various socio-economic variables. The objective of this experiment is to perform statistical data analysis to understand the distribution of data, identify patterns, detect outliers, and examine relationships between variables affecting income. Students will apply exploratory data analysis (EDA) techniques and build a Linear Regression model to predict income.

Tasks to be Performed:

• Load the dataset into Python environment.

• Identify numerical and categorical variables.

• Generate Histograms for all numerical columns.

• Create Boxplots to identify outliers in the dataset.

• Draw Violin plots to study data distribution and density.

• Plot Scatter graphs between Income and other variables to observe relationships.

• Fit Normal Distribution curves for each numerical feature and comment on distribution type.

• Split the dataset into training and testing sets.

• Apply Linear Regression for Income prediction.

• Evaluate the model using suitable performance metrics (e.g., R^2, error analysis).

Expected Outcome:

Students will gain hands-on experience in statistical data analysis, visualization techniques, understanding real-world data distribution, and developing a basic predictive model using Linear Regression.