Bias Detection Report

The dataset analyzed is the Statlog dataset, which contains information about loan applications.

The features examined are 'Status of existing checking account', 'Duration in month', 'Credit history', 'Purpose', 'Credit amount', 'Savings account/bonds', 'Present employment since', 'Installment rate in percentage of disposable income', 'Personal status and sex', 'Other debtors / guarantors', 'Present residence since', 'Property', 'Age in years', 'Other installment plans ', 'Housing', 'Number of existing credits at this bank', 'Job', 'Number of people being liable to provide maintenance for', 'Telephone', 'foreign worker', and 'Credit risk (Good/Bad)'

The type of bias detected is distribution bias, and the relevant feature is 'Purpose'

The tools used are the categorical_distribution_shannon_balance function to calculate the Shannon entropy and Balance metric for the loan purpose column, the get_reference_method_by_id function to get a method from the reference literature, the execute_python_code function to execute the method, and the plot_bar_chart function to visualize the distribution of the loan purpose column

The values obtained are a Shannon entropy of 2.666677759851852 and a Balance metric of 0.802749994485438 for the loan purpose column

The extent of the bias is moderate, as indicated by the Balance metric

Bar Chart of Purpose