

Loan Prediction Project - Notebook Report

Objective

The notebook is focused on exploratory data analysis (EDA) of a loan prediction dataset (train.csv). It is part of a Loan Approval Prediction project, often used in data science and machine learning tasks.

Key Operations Performed

- Data Loading: Reading the dataset using pandas.
- Initial Data Inspection: Using `.head()`, `.shape`, `.info()`, `.describe()`.
- EDA: Frequency counts, boxplots, and histograms.
- Data Cleaning: Filling missing values and handling categorical variables.
- Feature Engineering: Creating new features like `Total_Income`.
- Model Building: Logistic Regression, Decision Trees, Random Forest Classifier.
- Model Evaluation: Accuracy scores and Cross-validation.

Important Results

Logistic Regression and Random Forest models were evaluated and compared based on their accuracy scores. Cross-validation helped in checking model reliability.

Brief Analysis

The notebook has a good flow: EDA -> Cleaning -> Feature Engineering -> Model Building -> Evaluation. Visualizations are basic but effective. No markdown explanations or hyperparameter tuning were observed.

Suggestions for Improvement

- Add markdown cells explaining major steps.
- Perform correlation analysis.
- Do hyperparameter optimization (Grid Search/Randomized Search).
- Visualize feature importance.
- Compare models systematically using a table.