

Project Summary: Loan Eligibility Prediction

[\(Link to project\)](#)

1. Introduction:

The project aims to predict whether a person is eligible for a loan or not based on various factors such as gender, marital status, income, credit history, etc. This involves data preprocessing, exploratory data analysis (EDA), model training, and evaluation.

2. Libraries Used:

- Pandas: For data manipulation and analysis.
- NumPy: For numerical computations.
- Matplotlib and Seaborn: For data visualization.
- Scikit-learn: For machine learning models and evaluation.

3. Data Preprocessing:

- Loaded dataset from a CSV file.
- Checked for missing values and performed imputation for both numerical and categorical features.
- Converted categorical variables into numeric format using label encoding and one-hot encoding.
- Pre-processed the 'Dependents' column by removing the '+' sign and converting it to an integer.

4. Exploratory Data Analysis (EDA):

- Conducted EDA to understand the distribution of variables and their relationships with the target variable (Loan_Status).
- Visualized the distribution of categorical variables using count plots and numerical variables using histograms and boxplots.
- Analysed correlations between numerical variables using a correlation heatmap and pairplot.
- Identified conclusions based on the EDA findings, such as the distribution of loan status among different categories and correlations between variables.

5. Model Training:

- Split the dataset into training and testing sets.
- Trained multiple supervised learning models, including Logistic Regression, Random Forest, and Gradient Boosting.
- Evaluated model performance using accuracy scores.

6. Results:

- Logistic Regression Accuracy: 78.86%
- Random Forest Accuracy: 78.05%
- Gradient Boosting Accuracy: 76.42%
- Support Vector Machine (SVM) Accuracy: 79.46%
- K-Nearest Neighbours (KNN) Accuracy: 56.22%

7. Conclusion:

- Support Vector Machine (SVM) achieved the highest accuracy among the supervised learning models, followed by Logistic Regression and Random Forest.
- K-Nearest Neighbours (KNN) performed poorly compared to other models, indicating that it may not be suitable for this dataset.
- The project provides insights into factors influencing loan eligibility and demonstrates the application of machine learning techniques for predictive analytics in the financial domain.