

Loan Eligibility Classification Using Logistic Regression

Project Overview

This project focused on building a predictive model to assess loan eligibility using Logistic Regression, a supervised learning classification technique. The goal was to classify potential borrowers as eligible or ineligible, providing actionable insights for risk-aware lending decisions.

Data and Methodology

The dataset was preprocessed to handle missing values, encode categorical variables, and scale numerical features. Logistic Regression was selected for its interpretability and effectiveness in binary classification tasks. Model training included splitting the dataset into training and testing sets, followed by evaluation using standard classification metrics.

Model Performance Metrics

Accuracy: 0.8181 – Correct classification of over 81% of applicants.

Precision: 0.80 – 80% of approved applications were genuinely eligible.

Recall: 1.00 – All eligible applicants were correctly identified (no False Negatives).

F1-Score: 0.8888 – Balanced performance across precision and recall.

ROC-AUC: 0.606 – Measure of model's discriminative ability.

Confusion Matrix

	Predicted Eligible	Predicted Ineligible
Actual Eligible	56 True Positives	0 False Negatives
Actual Ineligible	14 False Positives	7 True Negatives

Insights and Interpretation

Perfect Recall (1.00): Ensures that all eligible borrowers are correctly approved, minimizing missed revenue opportunities.

Strong Precision (0.80): Indicates a manageable level of False Positives, maintaining controlled risk exposure.

The high F1-Score demonstrates that the model balances precision and recall effectively, making it suitable for automated lending decisions.

Project Impact

The model provides a reliable, data-driven approach for loan approval, reducing risk while ensuring no eligible applicants are overlooked. This approach can be extended and scaled for real-world lending platforms, offering a robust tool for decision-making in FinTech environments.