

LOAN APPROVAL CLASSIFICATION AUTOMATING CREDIT ASSESSMENT

LEVERAGING MACHINE LEARNING FOR EFFICIENT AND FAIR LOAN DECISIONS

2025



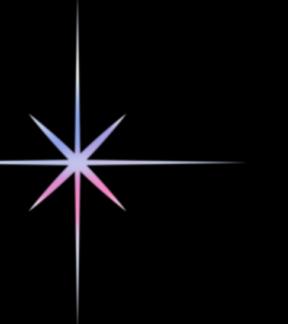
Problem Statement

- Loan approval decisions are critical for financial institutions but challenging due to:
- Financial losses from loan defaults.
- Missed opportunities from rejecting creditworthy applicants.
- Manual processes leading to inefficiencies.

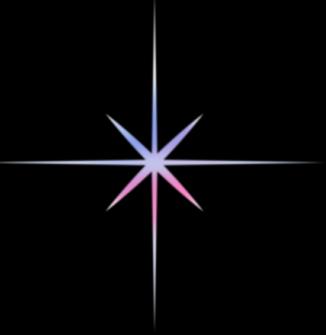
OBJECTIVE:

BUILD A CLASSIFICATION MODEL TO PREDICT LOAN APPROVAL (APPROVED/NOT APPROVED) TO STREAMLINE DECISIONS, MINIMIZE DEFAULTS, AND ENSURE FAIRNESS.

REDUCE LOSSES, FASTER APPROVALS, BETTER CUSTOMER EXPERIENCE.



Model Pipeline



DATA CLEANING

detecting and correcting (or removing) inaccurate, incomplete, or irrelevant data from a dataset.

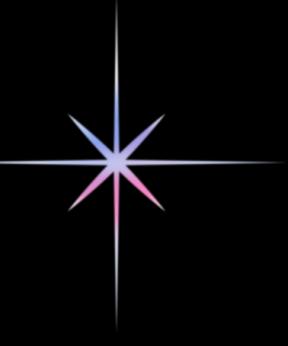
FEATURE ENGINEERING

creating new input features or modifying existing ones to improve a machine learning model's performance

MODEL SELECTION

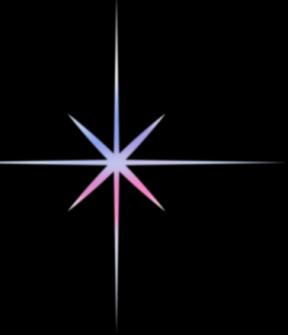
creating new input features or modifying existing ones to improve a machine learning model's performance

Insights from EDA



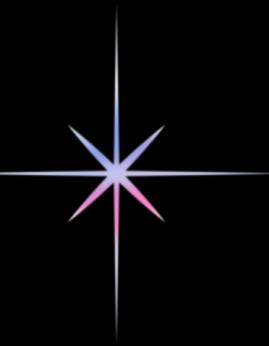
- Only 22% of the loans are getting approved of the total
- Applied Log transformation due to the skewness of the data as most of the features are right skewed
- Handled outliers as there are many unreal datas

Model Training and Tuning



- Algorithms Tested: Logistic Regression, Decision Tree, Random Forest, SVM.
- Pipeline: Preprocessing → SMOTE → Classifier.
- Hyperparameter Tuning: GridSearchCV with recall as the primary metric.

Model Performance Comparison



Model	Test Recall	Accuracy
Logistic Regression	0.9134	0.86
Decision Tree	0.8821	0.89
Random Forest	0.9035	0.89
SVM	0.9652	0.8

BEST MODEL DEEP DIVE

- Model:
SVM (C=0.1, gamma='scale', kernel='linear', probability=True)
- Performance:
Recall: 0.9652
F1-score: 0.68
Accuracy: 0.80
- Why Chosen:
Maximizes recall
Aligning with the goal of identifying creditworthy applicants.