

Loan Classification Project Report

Prepared by: Atharva Shinde

1. Introduction

This project aims to develop a machine learning model for classifying loan applications as either approved or rejected based on various financial and personal attributes of applicants.

2. Dataset Overview

- Total Records: 41188
- Total Features: 60
- Target Variable: Loan_Status_label (0 = Rejected, 1 = Approved)
- Sample of Dataset:

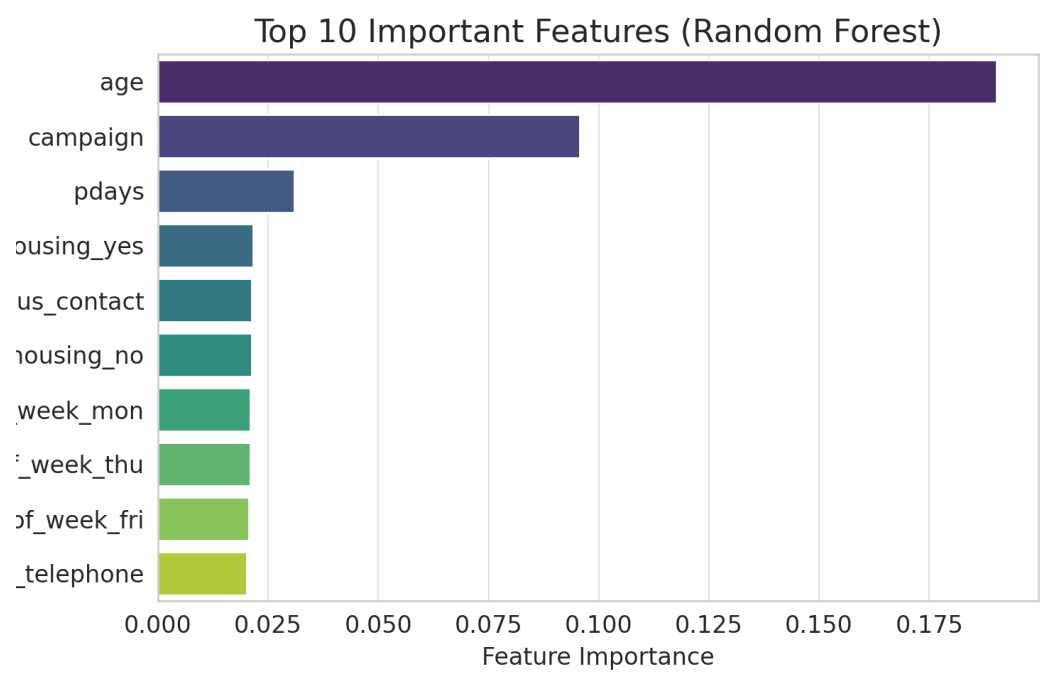
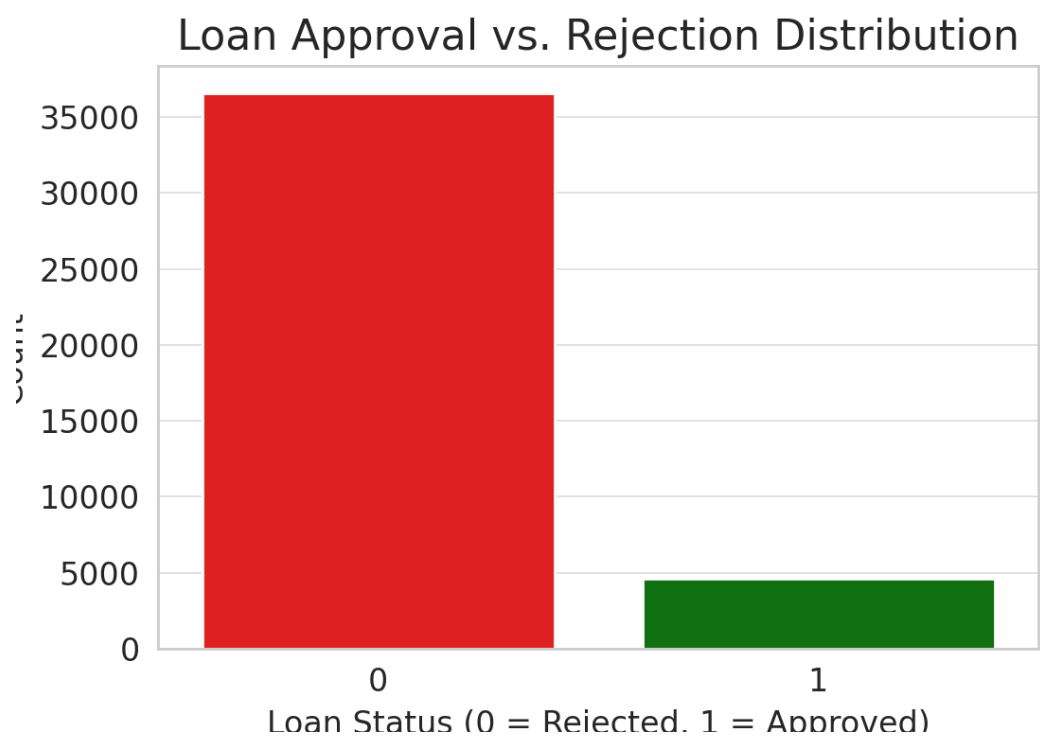
	age	campaign	pdays	previous	no_previous_contact	not_working	job_admin.	job_blue-collar	job_entrepreneur
0	56	1	999	0	1	0	0	1	0
1	57	1	999	0	1	0	0	0	0
2	37	1	999	0	1	0	0	0	0
3	40	1	999	0	1	0	1	0	0
4	56	1	999	0	1	0	0	0	0
5	45	1	999	0	1	0	0	0	0
6	59	1	999	0	1	0	1	0	0
7	41	1	999	0	1	0	0	1	0
8	24	1	999	0	1	0	0	0	0
9	25	1	999	0	1	0	0	0	0

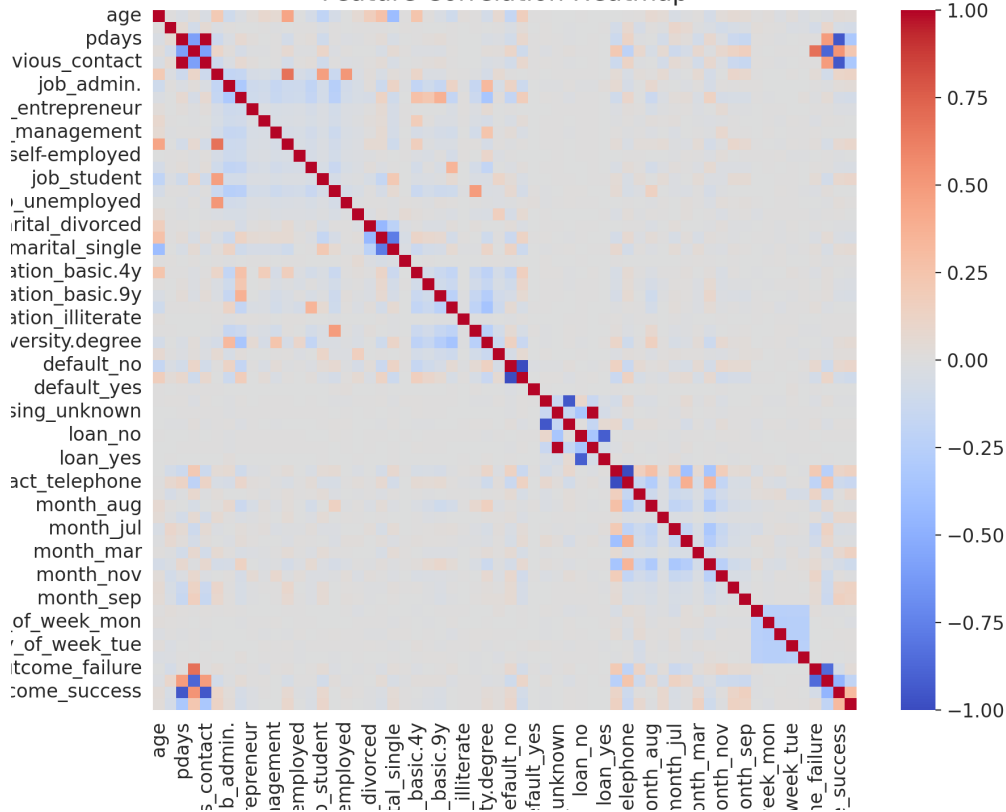
3. Data Summary

	count	mean	std	min	25%	50%	75%	max
age	41188.0	40.024060	10.421250	17.0	32.0	38.0	47.0	98.0
campaign	41188.0	2.567593	2.770014	1.0	1.0	2.0	3.0	56.0
pdays	41188.0	962.475454	186.910907	0.0	999.0	999.0	999.0	999.0
previous	41188.0	0.172963	0.494901	0.0	0.0	0.0	0.0	7.0
no_previous_contact	41188.0	0.963217	0.188230	0.0	1.0	1.0	1.0	1.0
not_working	41188.0	0.087623	0.282749	0.0	0.0	0.0	0.0	1.0
job_admin.	41188.0	0.253035	0.434756	0.0	0.0	0.0	1.0	1.0
job_blue-collar	41188.0	0.224677	0.417375	0.0	0.0	0.0	0.0	1.0
job_entrepreneur	41188.0	0.035350	0.184665	0.0	0.0	0.0	0.0	1.0
job_housemaid	41188.0	0.025736	0.158348	0.0	0.0	0.0	0.0	1.0
job_management	41188.0	0.070992	0.256814	0.0	0.0	0.0	0.0	1.0
job_retired	41188.0	0.041760	0.200042	0.0	0.0	0.0	0.0	1.0
job_self-employed	41188.0	0.034500	0.182513	0.0	0.0	0.0	0.0	1.0
job_services	41188.0	0.096363	0.295092	0.0	0.0	0.0	0.0	1.0
job_student	41188.0	0.021244	0.144199	0.0	0.0	0.0	0.0	1.0

job_technician	41188.0	0.163713	0.370019	0.0	0.0	0.0	0.0	1.0
job_unemployed	41188.0	0.024619	0.154962	0.0	0.0	0.0	0.0	1.0
job_unknown	41188.0	0.008012	0.089152	0.0	0.0	0.0	0.0	1.0
marital_divorced	41188.0	0.111974	0.315339	0.0	0.0	0.0	0.0	1.0
marital_married	41188.0	0.605225	0.488808	0.0	0.0	1.0	1.0	1.0
marital_single	41188.0	0.280859	0.449424	0.0	0.0	0.0	1.0	1.0
marital_unknown	41188.0	0.001942	0.044029	0.0	0.0	0.0	0.0	1.0
education_basic.4y	41188.0	0.101389	0.301846	0.0	0.0	0.0	0.0	1.0
education_basic.6y	41188.0	0.055647	0.229242	0.0	0.0	0.0	0.0	1.0
education_basic.9y	41188.0	0.146766	0.353877	0.0	0.0	0.0	0.0	1.0
education_high.school	41188.0	0.231014	0.421486	0.0	0.0	0.0	0.0	1.0
education_illiterate	41188.0	0.000437	0.020901	0.0	0.0	0.0	0.0	1.0
education_professional.course	41188.0	0.127294	0.333306	0.0	0.0	0.0	0.0	1.0
education_university.degree	41188.0	0.295426	0.456239	0.0	0.0	0.0	1.0	1.0

4. Data Visualization





5. Data Preprocessing

- Handled class imbalance using class weighting.
- Standardized numerical features using StandardScaler.
- Split data into training (80%) and testing (20%).

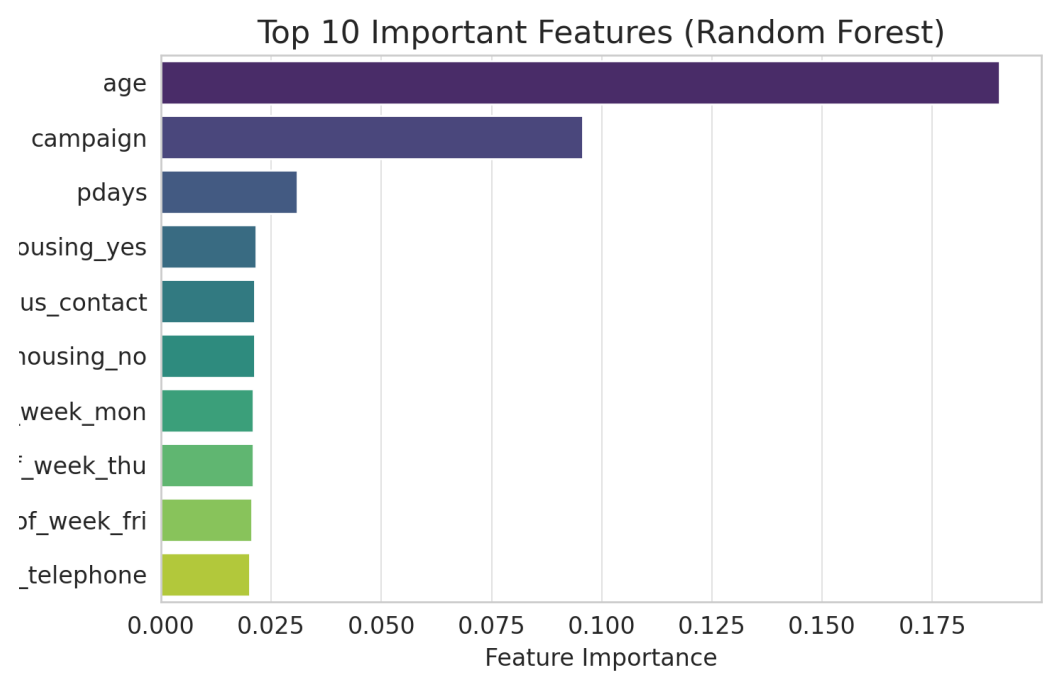
6. Model Training

- Algorithm Used: Random Forest Classifier
- Hyperparameters:
 - class_weight='balanced'
 - n_estimators=100
 - random_state=42

7. Model Evaluation

- Accuracy: 88.87%
- Precision, Recall, F1-score analyzed

8. Additional Visualizations



9. Conclusion

- The Random Forest model performed well but had challenges predicting approved loans.
- Future improvements could include oversampling (SMOTE), undersampling, or using models like XGBoost.