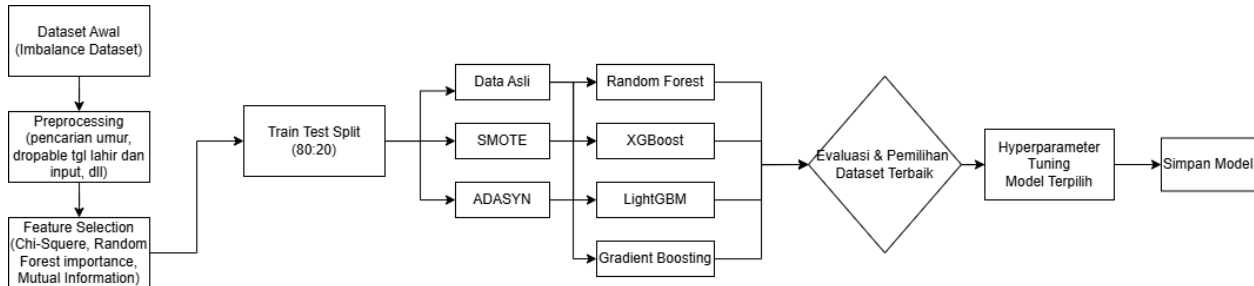


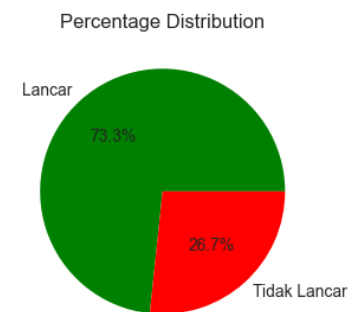
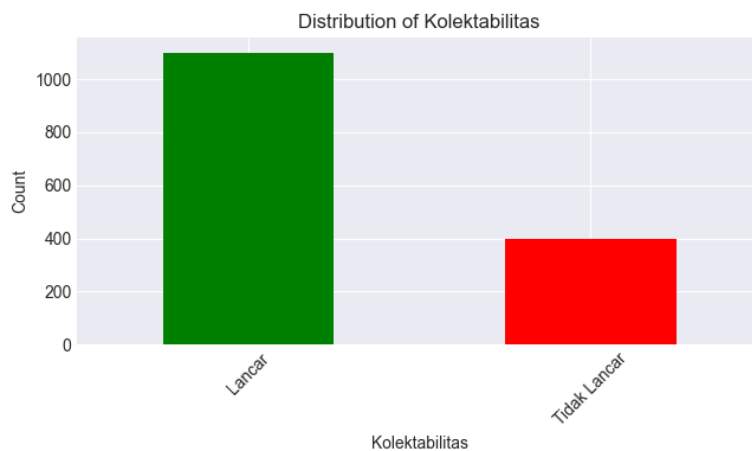
DOKUMENTASI PREDIKSI KREDIT MACET DENGAN ENSEMBLE LEARNING

1. Diagram Alur Analisis Model



2. Preprocessing & Penanganan Imbalance

2.1. LOADING DATA



- Dataset Original: 1500 baris × 24 kolom
- Target Variable: KOLEKTABILITAS (2 kategori)
 - Kelas 1: Lancar (1100 data)
 - Kelas 2: Tidak Lancar (400 data)

2.2

1.2. FEATURE SELECTION & REMOVAL

Kolom yang DIHAPUS (2 kolom):

1. Hasil Prescreening SLIK (*leakage_features*)
2. Hasil Prescreening (*leakage_features*)

Kolom yang DIPERTAHANKAN (13 kolom):

A. FITUR NUMERICAL (3 kolom):

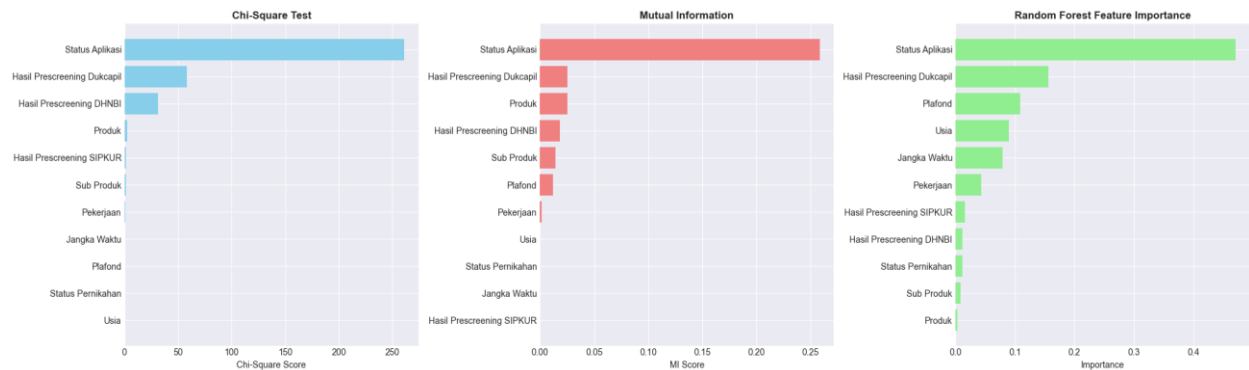
1. Pekerjaan
2. Usia
3. Status Pernikahan
4. Produk
5. Sub Produk

6. Plafond
7. Jangka Waktu
8. Hasil Prescreening SIPKUR
9. Hasil Prescreening Dukcapil
10. Hasil Prescreening DHNBI
11. Status Aplikasi

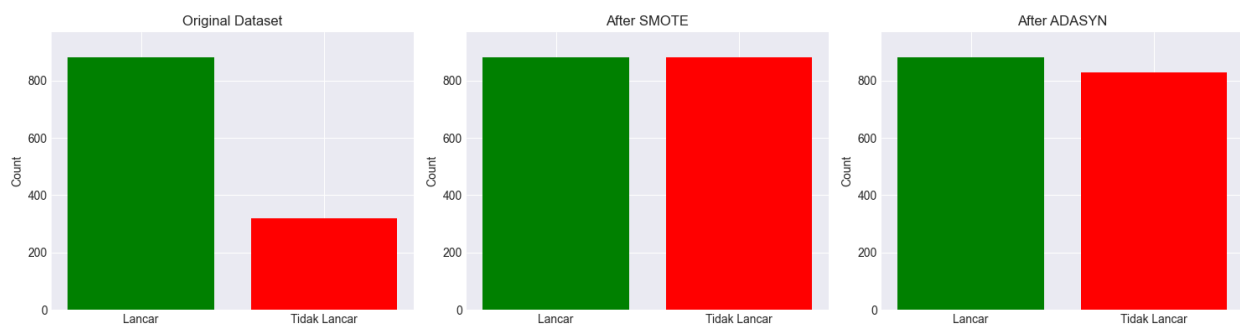
B. TARGET VARIABLE (1 kolom):

12. KOLEKTABILITAS : Target prediksi (2 kelas)

C. HASIL FEATURE SELECTION



D. HASIL OVERSAMPLING



1. ORIGINAL DATASET:

Training set distribution:

Lancar: 880 (73.33%)

Tidak Lancar: 320 (26.67%)

2. APPLYING SMOTE...

After SMOTE:

Lancar: 880 (50.00%)

Tidak Lancar: 880 (50.00%)

3. APPLYING ADASYN...

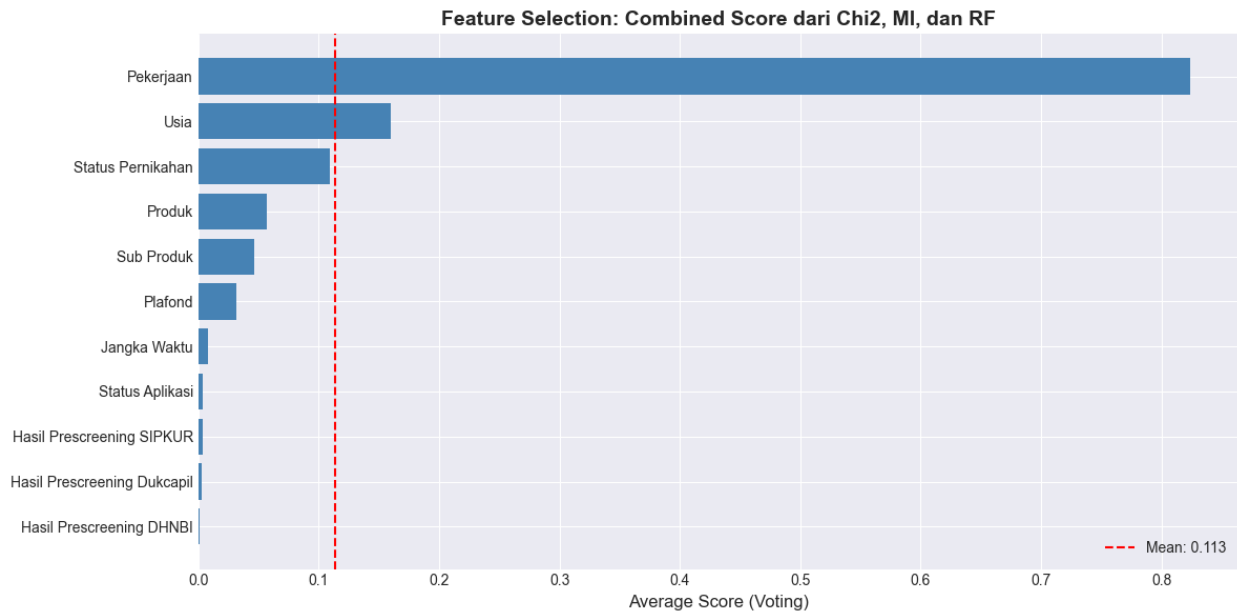
After ADASYN:

Lancar: 880 (51.49%)

Tidak Lancar: 829 (48.51%)

4. SUMMARY:

Original : 1200 samples
SMOTE : 1760 samples (+560)
ADASYN : 1709 samples (+509)
Test set : 300 samples (same for all)



SELECTED FEATURES: 10 dari 11

Features selected by RFE: 7

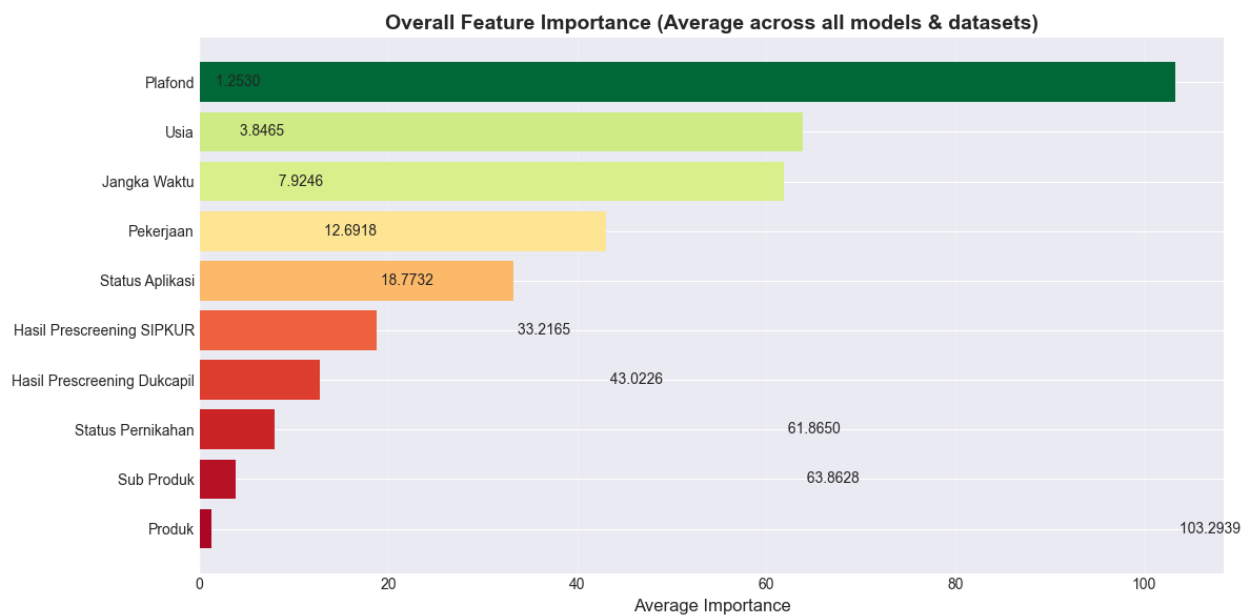
Features selected by Score: 7

Final union: 10

Final Selected Features (sorted by combined score):

1. Pekerjaan (Score: 0.8236) ✓ RFE ✓ Score
2. Usia (Score: 0.1596) ✓ RFE ✓ Score
3. Status Pernikahan (Score: 0.1090) ✓ Score
4. Produk (Score: 0.0567) ✓ Score
5. Sub Produk (Score: 0.0464) ✓ Score
6. Plafond (Score: 0.0313) ✓ RFE ✓ Score
7. Jangka Waktu (Score: 0.0077) ✓ RFE ✓ Score
8. Status Aplikasi (Score: 0.0040) ✓ RFE
9. Hasil Prescreening SIPKUR (Score: 0.0037) ✓ RFE
10. Hasil Prescreening Dukcapil (Score: 0.0029) ✓ RFE

Feature Importance



3. HASIL EVALUASI MODEL PER DATASET

COMPLETE COMPARISON: ALL MODELS x ALL DATASETS						
Dataset	Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Original	Random Forest	0.916667	0.789474	0.9375	0.857143	0.943977
Original	XGBoost	0.910000	0.784946	0.9125	0.843931	0.944830
Original	LightGBM	0.920000	0.797872	0.9375	0.862069	0.947386
Original	Gradient Boosting	0.916667	0.795699	0.9250	0.855491	0.947045
SMOTE	Random Forest	0.916667	0.789474	0.9375	0.857143	0.946136
SMOTE	XGBoost	0.920000	0.780000	0.9750	0.866667	0.949602
SMOTE	LightGBM	0.923333	0.787879	0.9750	0.871508	0.952330
SMOTE	Gradient Boosting	0.923333	0.800000	0.9500	0.868571	0.950455
ADASYN	Random Forest	0.916667	0.795699	0.9250	0.855491	0.948977
ADASYN	XGBoost	0.920000	0.785714	0.9625	0.865169	0.950057
ADASYN	LightGBM	0.923333	0.793814	0.9625	0.870056	0.950057
ADASYN	Gradient Boosting	0.923333	0.806452	0.9375	0.867052	0.946193

SUMMARY BY DATASET (Average Metrics):					
Dataset	Accuracy	Precision	Recall	F1-Score	ROC-AUC
ADASYN	0.920833	0.795420	0.946875	0.864442	0.948821
Original	0.915833	0.791998	0.928125	0.854658	0.945810
SMOTE	0.920833	0.789338	0.959375	0.865972	0.949631

SUMMARY BY MODEL (Average Metrics):					
Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Gradient Boosting	0.921111	0.800717	0.937500	0.863705	0.947898
LightGBM	0.922222	0.793189	0.958333	0.867878	0.949924
Random Forest	0.916667	0.791549	0.933333	0.856592	0.946364
XGBoost	0.916667	0.783554	0.950000	0.858589	0.948163

Evaluasi dengan mempertahankan kolom Hasil Prescreening SLIK, Hasil Prescreening (aneh bingits bu)

COMPLETE COMPARISON: ALL MODELS × ALL DATASETS							
Dataset	Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	
Original	Random Forest	1.0	1.0	1.0	1.0	1.0	1.0
Original	XGBoost	1.0	1.0	1.0	1.0	1.0	1.0
Original	LightGBM	1.0	1.0	1.0	1.0	1.0	1.0
Original	Gradient Boosting	1.0	1.0	1.0	1.0	1.0	1.0
SMOTE	Random Forest	1.0	1.0	1.0	1.0	1.0	1.0
	XGBoost	1.0	1.0	1.0	1.0	1.0	1.0
	LightGBM	1.0	1.0	1.0	1.0	1.0	1.0
	Gradient Boosting	1.0	1.0	1.0	1.0	1.0	1.0
ADASYN	Random Forest	1.0	1.0	1.0	1.0	1.0	1.0
ADASYN	XGBoost	1.0	1.0	1.0	1.0	1.0	1.0
ADASYN	LightGBM	1.0	1.0	1.0	1.0	1.0	1.0
ADASYN	Gradient Boosting	1.0	1.0	1.0	1.0	1.0	1.0
SUMMARY BY DATASET (Average Metrics):							
Dataset	Accuracy	Precision	Recall	F1-Score	ROC-AUC		
ADASYN	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Original	1.0	1.0	1.0	1.0	1.0	1.0	1.0
SMOTE	1.0	1.0	1.0	1.0	1.0	1.0	1.0
SUMMARY BY MODEL (Average Metrics):							
Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC		
Gradient Boosting	1.0	1.0	1.0	1.0	1.0	1.0	1.0
LightGBM	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Random Forest	1.0	1.0	1.0	1.0	1.0	1.0	1.0
XGBoost	1.0	1.0	1.0	1.0	1.0	1.0	1.0