

Final Assignment

ID/X Partners - Data Scientist

Presented by
Dipo Timur



Bogor



dipotimur89@gmail.com



www.linkedin.com/in/dipo-timur



<https://github.com/dipotr89-yo>IDX-Partners-Data-Scientist->

PROBLEM IDENTIFICATION

Bakground

kemampuan dalam memprediksi risiko gagal bayar menjadi sangat penting untuk menjaga stabilitas dan profitabilitas lembaga pemberi pinjaman. Data pinjaman tahun 2007–2014 memberikan gambaran mengenai berbagai karakteristik peminjam, jenis pinjaman, serta status pelunasan yang dapat dianalisis untuk memahami pola risiko tersebut. Dengan semakin banyaknya peminjam dan faktor yang memengaruhi keberhasilan pembayaran, penerapan machine learnin menjadi penting untuk meningkatkan akurasi pengambilan keputusan kredit. Oleh karena itu, dilakukan analisis data terhadap informasi peminjam dalam periode tersebut guna membangun model prediktif yang dapat membantu mengelola risiko pinjaman secara efektif.

PROBLEM IDENTIFICATION

Goals & Objective

- Memprediksi status pinjaman (lancar/gagal) berdasarkan data historis pinjaman.
- Memberikan insight bagi perusahaan finansial untuk mengambil keputusan pemberian pinjaman.



Analysis Approach

- Data Understanding: Melihat struktur dan statistik dekriptif
- Data Cleaning: Membersihkan data untuk keperluan analisis.
- Feature Engineering: Menyiapkan data untuk modeling.
- Modeling: Menggunakan beberapa algoritma machine learning.



Modelling & Evaluation

Modeling

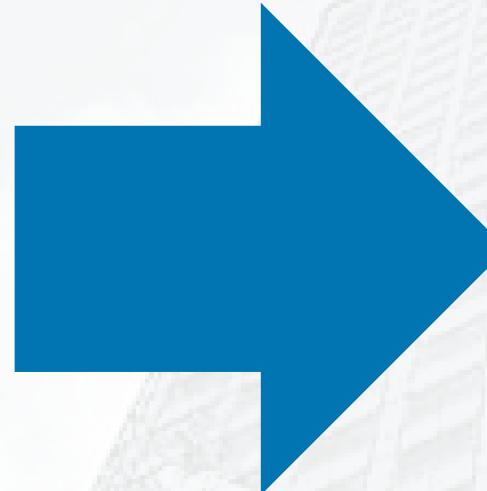
- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier

Evaluation Metrics:

- Accuracy Score
- Confusion Matrix
- Classification Report (precision, recall, F1-score)

Data Cleaning

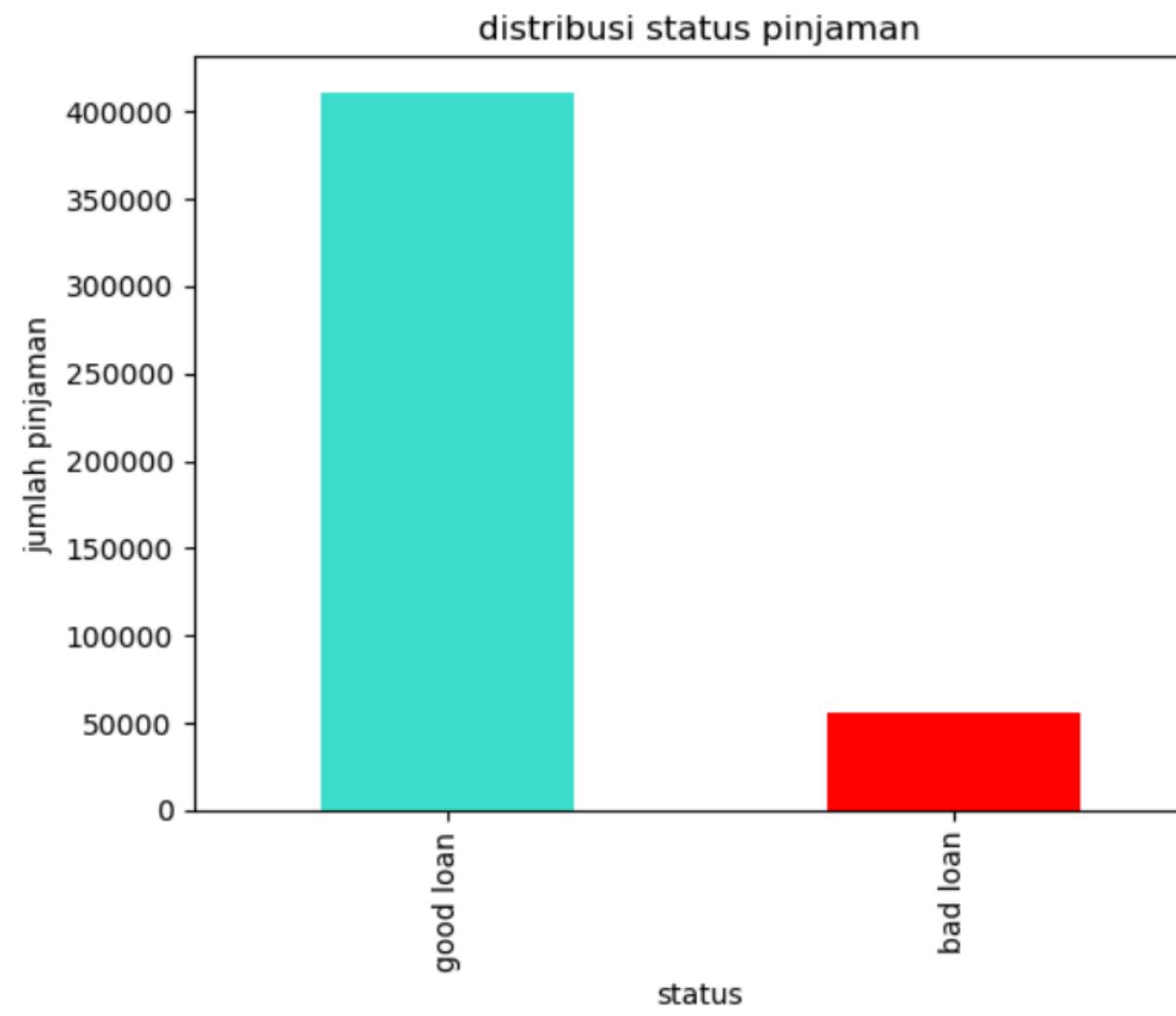
	Data Type	Missing Value	Percentage missing
Unnamed: 0	int64	0	0.000000
id	int64	0	0.000000
member_id	int64	0	0.000000
loan_amnt	int64	0	0.000000
funded_amnt	int64	0	0.000000
funded_amnt_inv	float64	0	0.000000
term	object	0	0.000000
int_rate	float64	0	0.000000
installment	float64	0	0.000000
grade	object	0	0.000000
sub_grade	object	0	0.000000
emp_title	object	27588	5.916553
emp_length	object	21008	4.505399
home_ownership	object	0	0.000000
annual_inc	float64	4	0.000858
verification_status	object	0	0.000000
issue_d	object	0	0.000000
loan_status	object	0	0.000000
pymnt_plan	object	0	0.000000
url	object	0	0.000000
purpose	object	0	0.000000
title	object	21	0.04504
zip_code	object	0	0.000000
addr_state	object	0	0.000000
dti	float64	0	0.000000
delinq_2yrs	float64	29	0.006219
earliest_cr_line	object	29	0.006219
inq_last_6mths	float64	29	0.006219
open_acc	float64	29	0.006219
pub_rec	float64	29	0.006219
revol_bal	int64	0	0.000000
revol_util	float64	340	0.072917
total_acc	float64	29	0.006219
initial_list_status	object	0	0.000000
out_prncp	float64	0	0.000000
out_prncp_inv	float64	0	0.000000
total_pymnt	float64	0	0.000000
total_pymnt_inv	float64	0	0.000000
total_rec_prncp	float64	0	0.000000
total_rec_int	float64	0	0.000000
total_rec_late_fee	float64	0	0.000000
recoveries	float64	0	0.000000
collection_recovery_fee	float64	0	0.000000
last_pymnt_d	object	376	0.080637
last_pymnt_amnt	float64	0	0.000000
last_credit_pull_d	object	42	0.009007
collections_12_mths_ex_med	float64	145	0.031097
policy_code	int64	0	0.000000
application_type	object	0	0.000000
acc_now_delinq	float64	29	0.006219
tot_coll_amt	float64	70276	15.071469
tot_cur_bal	float64	70276	15.071469
total_rev_hi_lim	float64	70276	15.071469



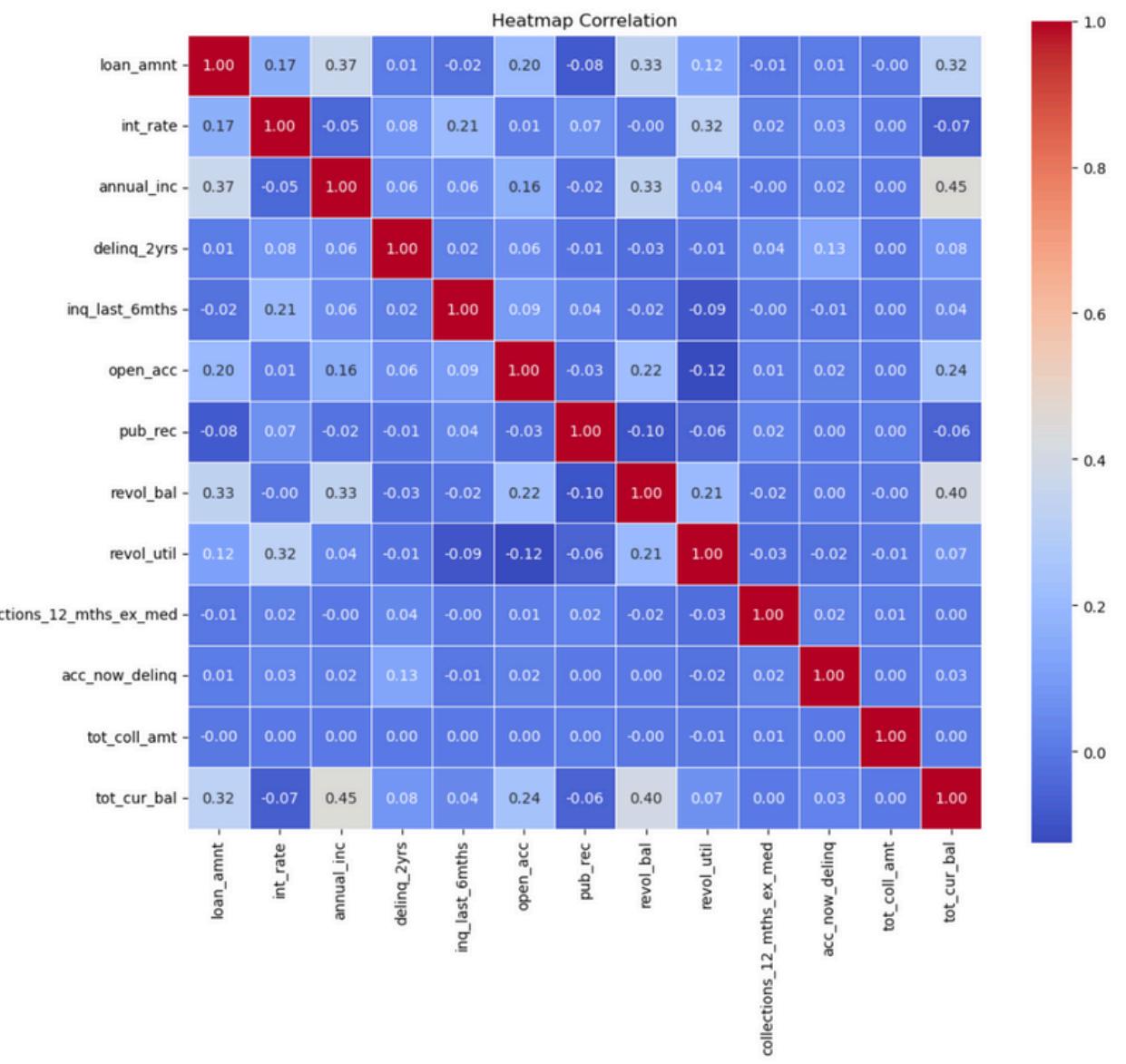
Unnamed: 0	0
id	0
member_id	0
loan_amnt	0
funded_amnt	0
funded_amnt_inv	0
term	0
int_rate	0
installment	0
grade	0
sub_grade	0
emp_title	0
emp_length	0
home_ownership	0
annual_inc	0
verification_status	0
issue_d	0
loan_status	0
pymnt_plan	0
url	0
purpose	0
title	0
zip_code	0
addr_state	0
dti	0
delinq_2yrs	0
earliest_cr_line	0
inq_last_6mths	0
open_acc	0
pub_rec	0
revol_bal	0
revol_util	0
total_acc	0
initial_list_status	0
out_prncp	0
out_prncp_inv	0
total_pymnt	0
total_pymnt_inv	0
total_rec_prncp	0
total_rec_int	0
total_rec_late_fee	0
recoveries	0
collection_recovery_fee	0
last_pymnt_d	0
last_pymnt_amnt	0
last_credit_pull_d	0
collections_12_mths_ex_med	0
policy_code	0
application_type	0
acc_now_delinq	0
tot_coll_amt	0
tot_cur_bal	0
total_rev_hi_lim	0
dtype:	int64

Feature Engineering

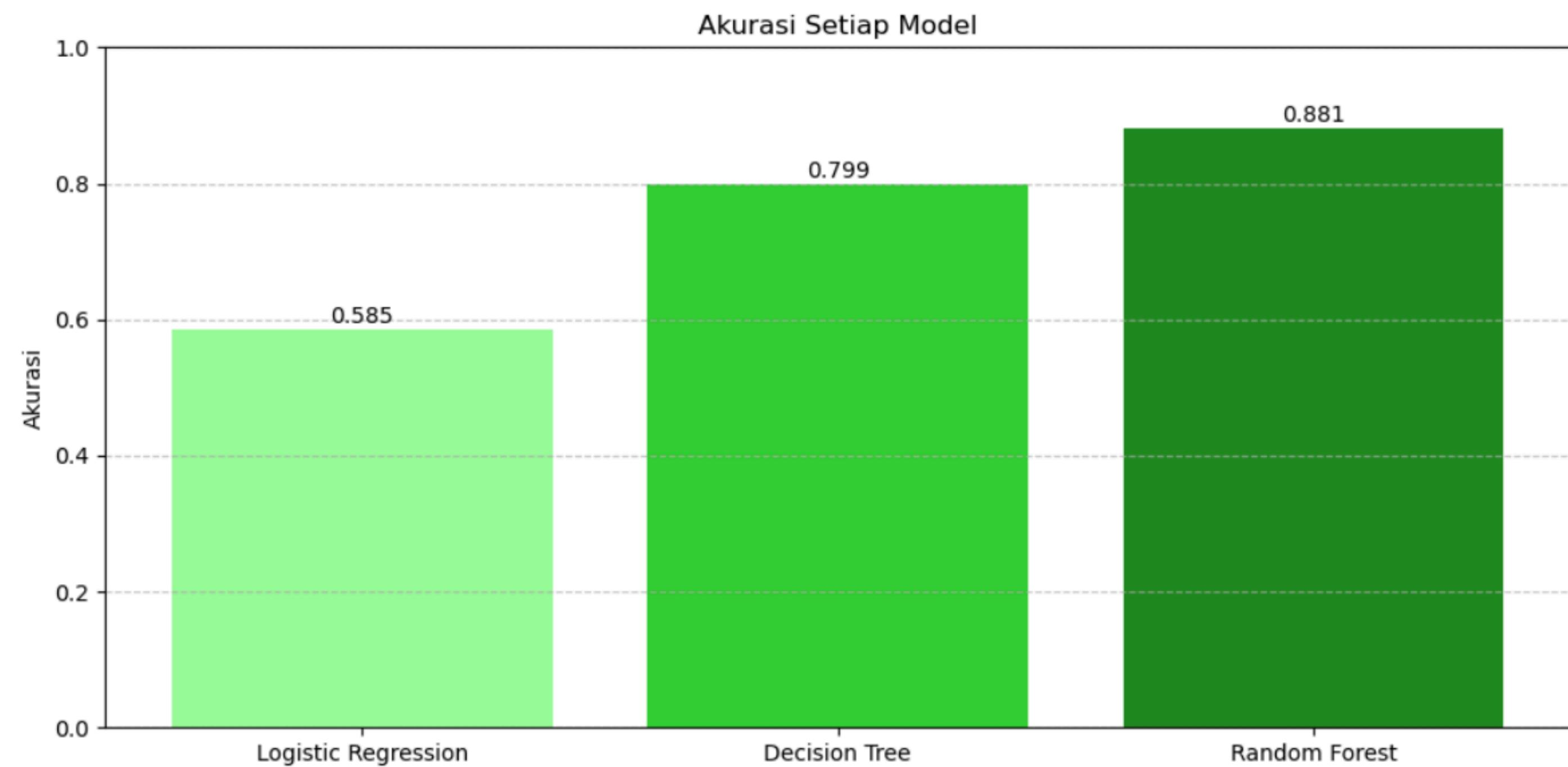
Variabel Dependen



Variabel Independen



Modelling



Conclusion

Random Forest Classifier terbukti memberikan hasil paling akurat dalam memprediksi status pembayaran pinjaman. Dibandingkan dengan model lain seperti Logistic Regression dan Decision Tree, Random Forest lebih unggul karena kemampuannya menangkap pola data yang kompleks dengan cara menggabungkan banyak pohon keputusan untuk menghasilkan prediksi yang lebih konsisten dan andal.

Dalam konteks dataset , model Random Forest mampu memanfaatkan berbagai informasi penting untuk mengenali faktor-faktor yang berperan besar dalam menentukan apakah seseorang akan mampu melunasi pinjamannya atau tidak. Artinya, model ini dapat membantu membedakan dengan lebih akurat mana peminjam yang berisiko dan mana yang tidak.

Thank You

Link Video :

<https://drive.google.com/drive/folders/1hgdnA8KBsRrT5n4RtQRnAC9hUEkt-nR3?usp=sharing>

