CREDIT RISK MODELLING WITH MACHINE LEARNING

Fidya Almira Suheri

OBJECTIVE

create a model to determine or predict credit risk based on the classification of good and bad loans

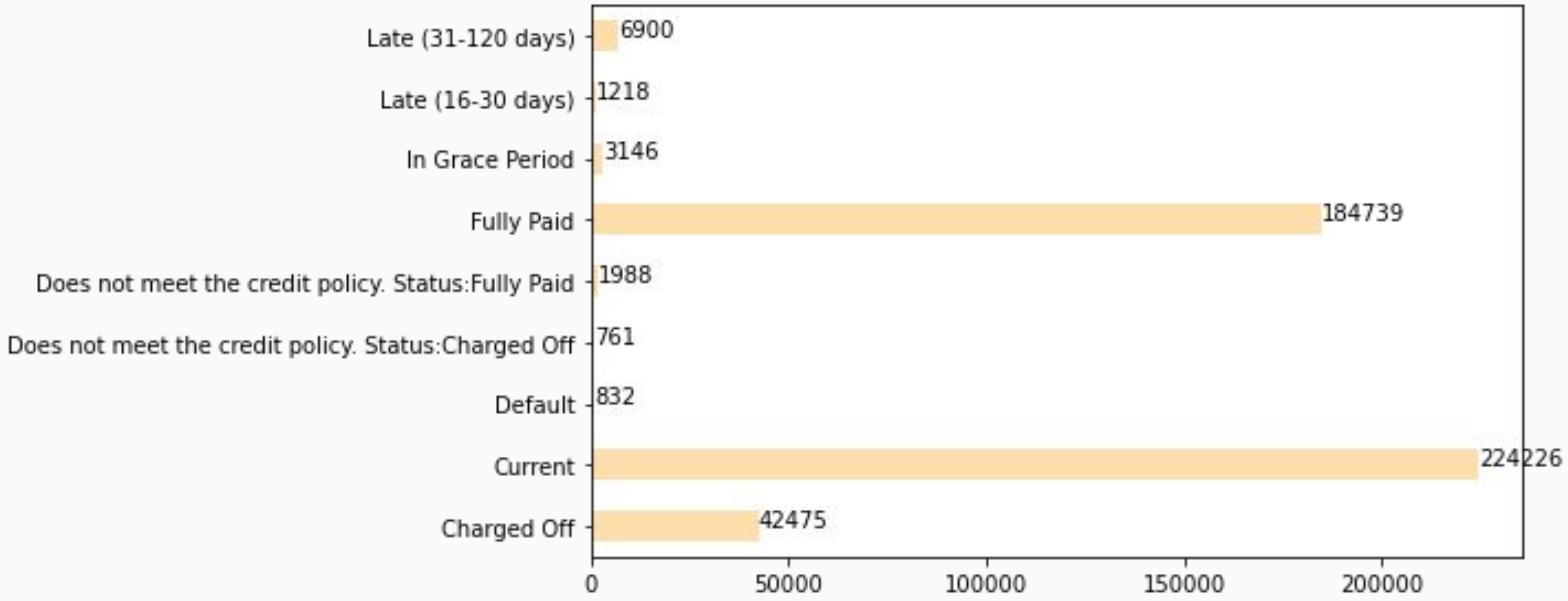
01
TARGET

02
MODELLING

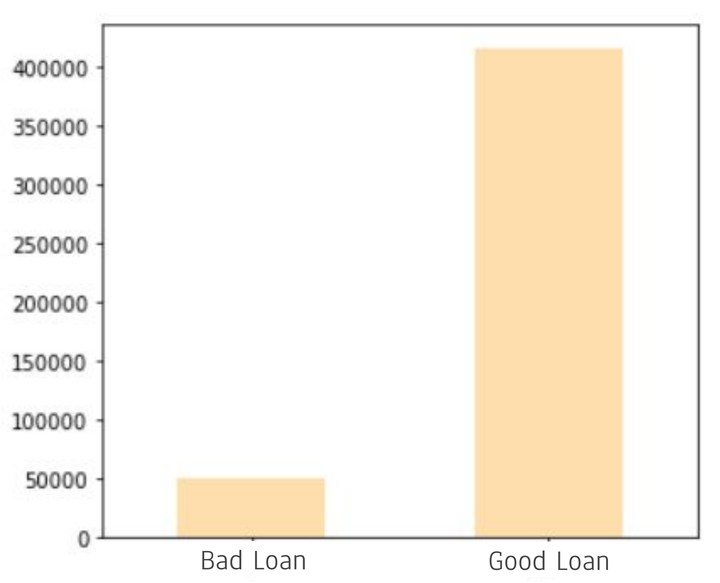
03
FEATURES

04
EVALUATION

TARGET/LABEL/CLASS



LOAN STATUS



Status loan dibagi menjadi 2 tipe:

Good Loan

- Current
- Fully Paid
- In Grace Period
- Late (16-30) days
- Does not meet the credit policy. Status: Fully Paid

Bad Loan

- Charged Off
- Default'
- Late (31-120 days)
- Does not meet the credit policy. Status:Charged Off'

PERSENTASE FEATURES DENGAN MISSING VALUE > 70%

Features	Persentase	Features	Persentase
desc	72.9815%	open_il_6m	100%
mths_since_last_record	86.5666%	open_il_12m	100%
mths_since_last_major_derog	78.7739%	open_il_24m	100%
annual_inc_joint	100%	mths_since_rcnt_il	100%
dti_joint	100%	total_bal_il	100%
verification_status_joint	100%	il_util	100%
open_acc_6m	100%	open_rv_12m	100%
open_rv_24m	100%	max_bal_bc	100%
all_util	100%	total_cu_tl	100%
inq_fi	100%	inq_last_12m	100%

FEATURE

Features	Description
id	A unique LC assigned ID for the loan listing.
member_id	A unique LC assigned Id for the borrower member.
url	URL for the LC page with listing data.
zip_code	The first 3 numbers of the zip code provided by the borrower in the loan application.
policy_code	publicly available policy_code=1 new products not publicly available policy_code=2
emp_title	The job title supplied by the Borrower when applying for the loan.
title	The loan title provided by the borrower

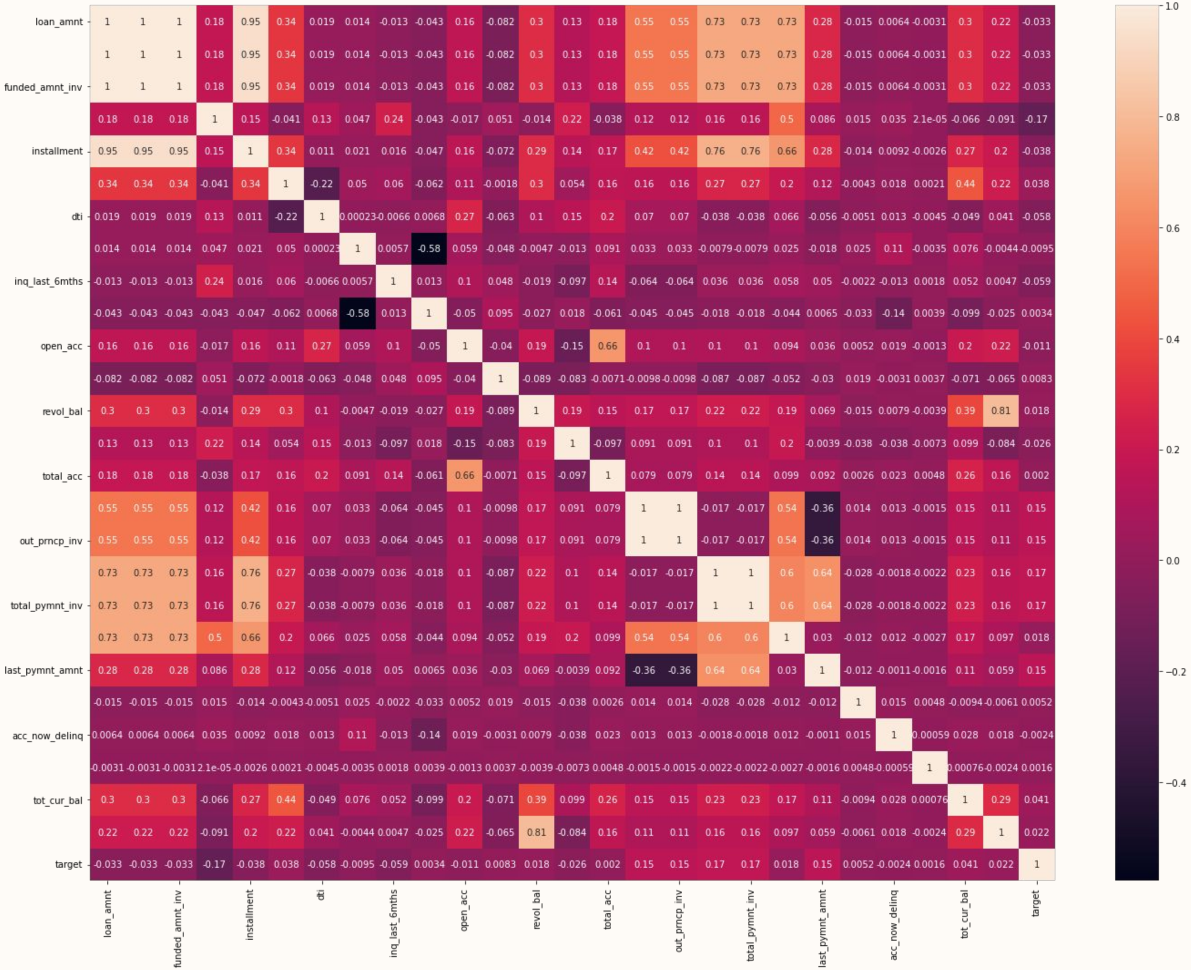
*Features yang dapat diabaikan

FEATURE

Features	Description
collection_recovery_fee	post charge off collection fee
next_pymnt_d	Next scheduled payment date
recoveries	post charge off gross recovery
total_rec_prncp	Principal received to date
total_rec_late_fee	Late fees received to date
sub_grade	LC assigned loan subgrade

*Features yang dapat diabaikan

KORELASI ANTAR FEATURES



	loan_amnt		funded_amnt_inv		installment		dti		inq_last_6mths		open_acc		revol_bal		total_acc		out_prncp_inv		total_pymnt_inv		last_pymnt_amnt		acc_now_delinq		tot_cur_bal		target
loan_amnt	1	1	1	0.18	0.95	0.34	0.019	0.014	-0.013	-0.043	0.16	-0.082	0.3	0.13	0.18	0.55	0.55	0.73	0.73	0.73	0.28	-0.015	0.0064	-0.0031	0.3	0.22	-0.033
	1	1	1	0.18	0.95	0.34	0.019	0.014	-0.013	-0.043	0.16	-0.082	0.3	0.13	0.18	0.55	0.55	0.73	0.73	0.73	0.28	-0.015	0.0064	-0.0031	0.3	0.22	-0.033
funded_amnt_inv	1	1	1	0.18	0.95	0.34	0.019	0.014	-0.013	-0.043	0.16	-0.082	0.3	0.13	0.18	0.55	0.55	0.73	0.73	0.73	0.28	-0.015	0.0064	-0.0031	0.3	0.22	-0.033
	0.18	0.18	0.18	1	0.15	-0.041	0.13	0.047	0.24	-0.043	-0.017	0.051	-0.014	0.22	-0.038	0.12	0.12	0.16	0.16	0.5	0.086	0.015	0.035	2.1e-05	-0.066	-0.091	-0.17
installment	0.95	0.95	0.95	0.15	1	0.34	0.011	0.021	0.016	-0.047	0.16	-0.072	0.29	0.14	0.17	0.42	0.42	0.76	0.76	0.66	0.28	-0.014	0.0092	-0.0026	0.27	0.2	-0.038
	0.34	0.34	0.34	-0.041	0.34	1	-0.22	0.05	0.06	-0.062	0.11	-0.0018	0.3	0.054	0.16	0.16	0.16	0.27	0.27	0.2	0.12	-0.0043	0.018	0.0021	0.44	0.22	0.038
dti	0.019	0.019	0.019	0.13	0.011	-0.22	1	0.00023	-0.0066	0.0068	0.27	-0.063	0.1	0.15	0.2	0.07	0.07	-0.038	-0.038	0.066	-0.056	-0.0051	0.013	-0.0045	-0.049	0.041	-0.058
	0.014	0.014	0.014	0.047	0.021	0.05	0.00023	1	0.0057	-0.58	0.059	-0.048	-0.0047	-0.013	0.091	0.033	0.033	-0.0079	-0.0079	0.025	-0.018	0.025	0.11	-0.0035	0.076	-0.0044	-0.0095
inq_last_6mths	-0.013	-0.013	-0.013	0.24	0.016	0.06	-0.0066	0.0057	1	0.013	0.1	0.048	-0.019	-0.097	0.14	-0.064	-0.064	0.036	0.036	0.058	0.05	-0.0022	-0.013	0.0018	0.052	0.0047	-0.059
	-0.043	-0.043	-0.043	-0.043	-0.047	-0.062	0.0068	-0.58	0.013	1	-0.05	0.095	-0.027	0.018	-0.061	-0.045	-0.045	-0.018	-0.018	-0.044	0.0065	-0.033	-0.14	0.0039	-0.099	-0.025	0.0034
open_acc	0.16	0.16	0.16	-0.017	0.16	0.11	0.27	0.059	0.1	-0.05	1	-0.04	0.19	-0.15	0.66	0.1	0.1	0.1	0.1	0.094	0.036	0.0052	0.019	-0.0013	0.2	0.22	-0.011
	-0.082	-0.082	-0.082	0.051	-0.072	-0.0018	-0.063	-0.048	0.048	0.095	-0.04	1	-0.089	-0.083	-0.0071	-0.0098	-0.0098	-0.087	-0.087	-0.052	-0.03	0.019	-0.0031	0.0037	-0.071	-0.065	0.0083
revol_bal	0.3	0.3	0.3	-0.014	0.29	0.3	0.1	-0.0047	-0.019	-0.027	0.19	-0.089	1	0.19	0.15	0.17	0.17	0.22	0.22	0.19	0.069	-0.015	0.0079	-0.0039	0.39	0.81	0.018
	0.13	0.13	0.13	0.22	0.14	0.054	0.15	-0.013	-0.097	0.018	-0.15	-0.083	0.19	1	-0.097	0.091	0.091	0.1	0.1	0.2	-0.0039	-0.038	-0.038	-0.0073	0.099	-0.084	-0.026
total_acc	0.18	0.18	0.18	-0.038	0.17	0.16	0.2	0.091	0.14	-0.061	0.66	-0.0071	0.15	-0.097	1	0.079	0.079	0.14	0.14	0.099	0.092	0.0026	0.023	0.0048	0.26	0.16	0.002
	0.55	0.55	0.55	0.12	0.42	0.16	0.07	0.033	-0.064	-0.045	0.1	-0.0098	0.17	0.091	0.079	1	1	-0.017	-0.017	0.54	-0.36	0.014	0.013	-0.0015	0.15	0.11	0.15
out_prncp_inv	0.55	0.55	0.55	0.12	0.42	0.16	0.07	0.033	-0.064	-0.045	0.1	-0.0098	0.17	0.091	0.079	1	1	-0.017	-0.017	0.54	-0.36	0.014	0.013	-0.0015	0.15	0.11	0.15
	0.73	0.73	0.73	0.16	0.76	0.27	-0.038	-0.0079	0.036	-0.018	0.1	-0.087	0.22	0.1	0.14	-0.017	-0.017	1	1	0.6	0.64	-0.028	-0.0018	-0.0022	0.23	0.16	0.17
total_pymnt_inv	0.73	0.73	0.73	0.16	0.76	0.27	-0.038	-0.0079	0.036	-0.018	0.1	-0.087	0.22	0.1	0.14	-0.017	-0.017	1	1	0.6	0.64	-0.028	-0.0018	-0.0022	0.23	0.16	0.17
	0.73	0.73	0.73	0.5	0.66	0.2	0.066	0.025	0.058	-0.044	0.094	-0.052	0.19	0.2	0.099	0.54	0.54	0.6	0.6	1	0.03	-0.012	0.012	-0.0027	0.17	0.097	0.018
last_pymnt_amnt	0.28	0.28	0.28	0.086	0.28	0.12	-0.056	-0.018	0.05	0.0065	0.036	-0.03	0.069	-0.0039	0.092	-0.36	-0.36	0.64	0.64	0.03	1	-0.012	-0.0011	-0.0016	0.11	0.059	0.15
	-0.015	-0.015	-0.015	0.015	-0.014	-0.0043	-0.0051	0.025	-0.0022	-0.033	0.0052	0.019	-0.015	-0.038	0.0026	0.014	0.014	-0.028	-0.028	-0.012	-0.012	1	0.015	0.0048	-0.0094	-0.0061	0.0052
acc_now_delinq	0.0064	0.0064	0.0064	0.035	0.0092	0.018	0.013	0.11	-0.013	-0.14	0.019	-0.0031	0.0079	-0.038	0.023	0.013	0.013	-0.0018	-0.0018	0.012	-0.0011	0.015	1	0.00059	0.028	0.018	-0.0024
	-0.0031	-0.0031	-0.0031	2.1e-05	-0.0026	0.0021	-0.0045	-0.0035	0.0018	0.0039	-0.0013	0.0037	-0.0039	-0.0073	0.0048	-0.0015	-0.0015	-0.0022	-0.0022	-0.0027	-0.0016	0.0048	-0.00059	1	0.00076	-0.0024	0.0016
tot_cur_bal	0.3	0.3	0.3	-0.066	0.27	0.44	-0.049	0.076	0.052	-0.099	0.2	-0.071	0.39	0.099	0.26	0.15	0.15	0.23	0.23	0.17	0.11	-0.0094	0.028	0.00076	1	0.29	0.041
	0.22	0.22	0.22	-0.091	0.2	0.22	0.041	-0.0044	0.0047	-0.025	0.22	-0.065	0.81	-0.084	0.16	0.11	0.11	0.16	0.16	0.097	0.059	-0.0061	0.018	-0.0024	0.29	1	0.022
target	-0.033	-0.033	-0.033	-0.17	-0.038	0.038	-0.058	-0.0095	-0.059	0.0034	-0.011	0.0083	0.018	-0.026	0.002	0.15	0.15	0.17	0.17	0.018	0.15	0.0052	-0.0024	0.0016	0.041	0.022	1

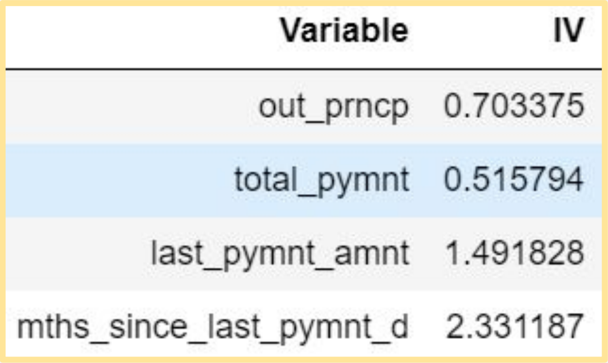
INFORMATION VALUE

$$IV = \sum (\text{Event\%} - \text{Non Event\%}) * \ln\left(\frac{\text{Event\%}}{\text{Non Event\%}}\right)$$

$$IV = \sum (\text{Event\%} - \text{Non Event\%}) * (\text{WOE})$$

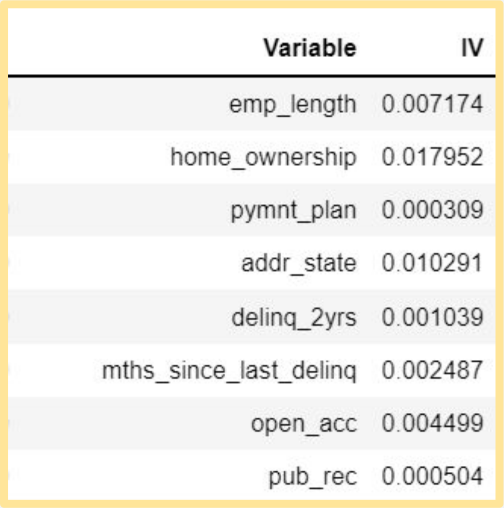
INFORMATION VALUE

Information Value>0.5

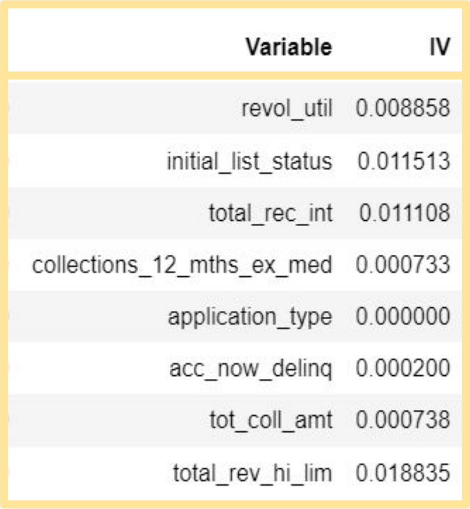


Variable	IV
out_prncp	0.703375
total_pymnt	0.515794
last_pymnt_amnt	1.491828
mths_since_last_pymnt_d	2.331187

Information Value<0.02



Variable	IV
emp_length	0.007174
home_ownership	0.017952
pymnt_plan	0.000309
addr_state	0.010291
delinq_2yrs	0.001039
mths_since_last_delinq	0.002487
open_acc	0.004499
pub_rec	0.000504



Variable	IV
revol_util	0.008858
initial_list_status	0.011513
total_rec_int	0.011108
collections_12_mths_ex_med	0.000733
application_type	0.000000
acc_now_delinq	0.000200
tot_coll_amt	0.000738
total_rev_hi_lim	0.018835

FEATURES

Kesimpulan

- Terdapat 75 features pada data asli dengan dengan 40 feature yang memiliki *Missing Value.*
- Feature yang memiliki 70% *missing value* akan dihapus karena akan menimbulkan ketidakakuratan ketika pemodelan.
- Feature yang merupakan data kejadian yang terjadi dimasa depan akan dihapus karena tidak memiliki pengaruh dalam pemodelan
- Feature yang memiliki korelasi yang sangat kuat dengan feature lain akan dihapus karena akan menghasilkan model yang kurang maksimal pada *logistic regression*
- Feature yang bertipe kategorikal akan dikonversi menjadi numerik disesuaikan dengan data yang terkait.
- Feature yang memiliki *Information Value* jauh dibawah 0.2 dan diatas 0.5 akan dihapus karena tidak dapat digunakan untuk pemodelan.

MODELLING

TRAIN TEST SPLIT :

Data dibagi menjadi 2 bagian, yaitu:

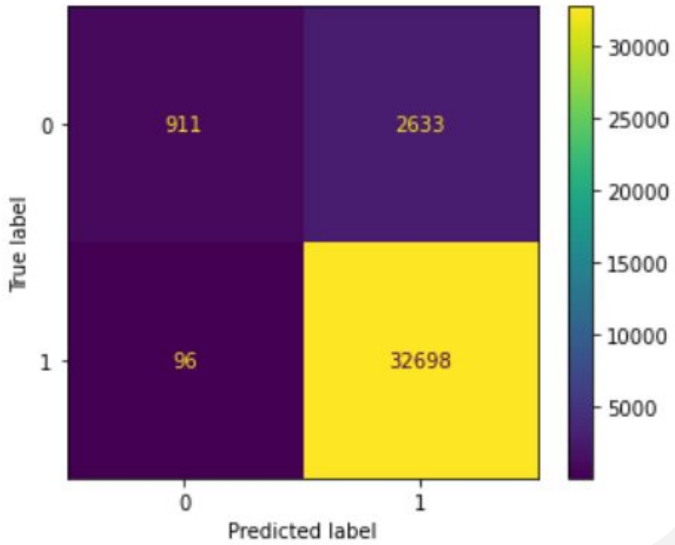
- Training Dataset = 80%
- Test Dataset = 20%

MODEL MACHINE LEARNING:

Classification : Logistic Regression

EVALUATION

Confusion Matrix



Classification Report

	Precision	Recall	f1-score	support
0	0.90	0.26	0.40	3544
1	0.93	1	0.96	32794
accuracy			0.92	36338
Macro avg	0.92	0.63	0.68	36338
Weigh avg	0.92	0.92	0.91	36338

Training Accuracy : 0.9262390951371879
Testing Accuracy : 0.9248995541857009