Classification: Non Performing Loan (NPL) Credit Scoring

Capstone - Deschie Tri Aksara

Problem

- Banyak faktor yang dipertimbangkan analis kredit sebelum memberikan rekomendasi menerima atau menolak sebuah pengajuan kredit
- Bank A ingin membuat sistem penilaian kelayakan kredit secara otomatis berdasarkan data yang mereka miliki.

Goal

- Melakukan analisis dan prediksi kredit macet berdasarkan variabel yang ada.
- Metrics yang dipertimbangkan untuk credit scoring yaitu recall



Sumber: NFCC.org

Data Description

- 1. X: Nomor
- **2. jumlah_kartu**: Jumlah kartu aktif yang dimiliki pelanggan
- 3. outstanding: Total saldo pemakaian kartu kredit
- **4. limit_kredit**: Jumlah maksimum limit kredit yang dapat digunakan
- 5. tagihan: Jumlah tagihan pada bulan terakhir
- **6. total_pemakaian_tunai**: Total pemakaian transaksi tunai pada bulan terakhir (cash advance)
- 7. total_pemakaian_retail: Total pemakaian transaksi retail pada bulan terakhir
- 8. sisa_tagihan_tidak_terbayar: Jumlah tagihan tidak terbayar pada bulan terakhir
- 9. rasio_pembayaran: Rasio perbandingan jumlah yang dibayar dengan tagihan pada bulan terakhir
- 10. persentasi overlimit: Presentasi overlimit
- **11.** rasio_pembayaran_3bulan: Rata-rata rasio pembayaran 3 bulan terakhir
- 12. rasio_pembayaran_6bulan: Rata-rata rasio pembayaran 3 bulan sebelum 3 bulan terakhir
- 13. skor_delikuensi: Kecenderungan pembayaran menunggak selama 12 bulan ke depan (skor rendah berarti buruk)
- **14. flag_kredit_macet**: Flag yang menandakan apakah kartu kredit macet atau tidak
- **15. jumlah_tahun_sejak_pembukaan_kredit**: Jumlah tahun sejak pembuatan kartu kredit pertama kali

Data Description

- **16. total_pemakaian**: Total pemakaian kartu kredit (tunai dan retail) pada bulan terakhir
- 17. sisa_tagihan_per_jumlah_kartu: Jumlah tagihan yang tidak terbayar pada bulan terakhir dibagi jumlah kartu aktif
- 18. sisa_tagihan_per_limit: Jumlah tagihan yang tidak terbayar pada bulan terakhir dibagi total limit
- 19. total_pemakaian_per_limit: Total pemakaian kartu kredit (tunai dan retail) pada bulan terakhir dibagi total limit
- 20. pemakaian_3bln_per_limit: Rata-rata pemakaian selama 3 bulan terakhir dibagi total limit
- 21. pemakaian_6bln_per_limit: Rata-rata pemakaian selama 3 bulan sebelum 3 bulan terakhir dibagi total limit
- 22. utilisasi 3bulan: Utilisasi kartu kredit selama 3 bulan terakhir
- 23. utilisasi_6bulan: Utilisasi kartu kredit selama 3 bulan sebelum 3 bulan terakhir

Response Variable / Target

flag_kredit_macet

0: Tidak Macet

1: Macet

Data Processing Schema

Load & read dataset

Exploratory
Data Analysis

Preprocessing

Feature
Engineering

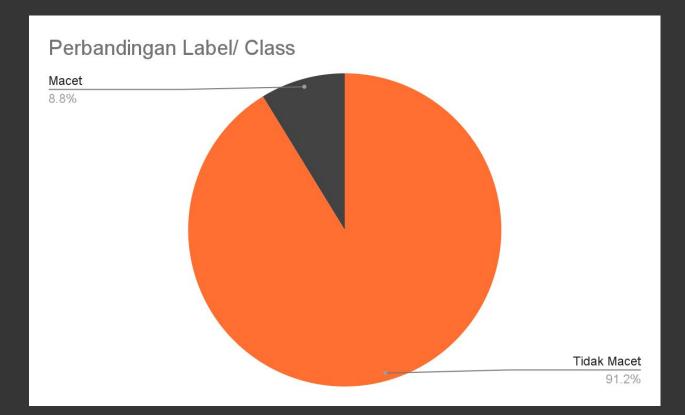
Modeling

Load and Predict
New Data

Conclusion and
Recommendation

- 1. Logistic Regression
- 2. Random Forest Classifier
- 3. Gradient Boosting Classifier

Exploratory Data Analysis

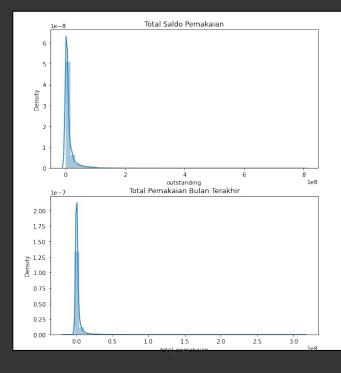


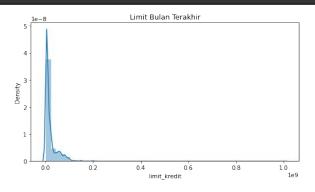


Proporsi data tidak seimbang sehingga akan dilakukan resampling dengan menggunakan SMOTE.

Analisa Deskriptif

- Jumlah kartu terbanyak adalah 16 kartu.
- Dari beberapa kolom seperti total_pemakaian, rasio_pembayaran, dan total_pemakaian_per_limit, nilai minimum bernilai negatif.







Outlier

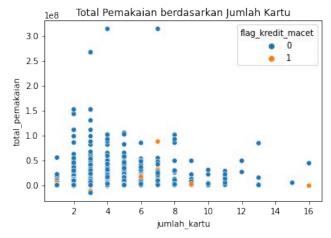
Terdapat perbedaan yang signifikan antara mean dan median dari kolom outstanding, limit_kredit, dan total_pemakaian yang mengindikasikan outlier atau skewed distribution.

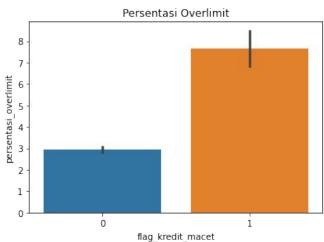


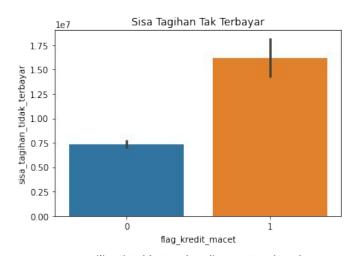
Exploratory Data Analysis

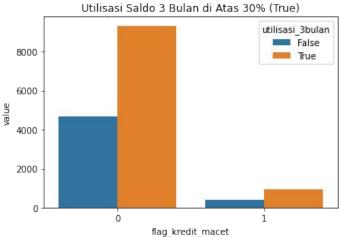
- 1. Apakah semakin banyak jumlah kartu semakin besar total saldo yang dipakai?
- 2. Apakah sisa tagihan tidak terbayar yang tinggi mengindikasikan kredit macet?
- 3. Bagaimana dengan persentasi overlimit yang besar terhadap flag kredit macet?
- 4. Apakah nasabah yang menggunakan saldo kreditnya di atas 30% (utilitas kredit > 30%) cenderung terindikasi kredit macet?

Exploratory Data Analysis









	Variabel	Jumlah Missing Value
	kode_cabang	100
drop kolom	lainnya	0

total_pemakaian_per_limit	-0.116207				
pemakaian_3bln_per_limit	-0.067138				
total_pemakaian_retail	-0.049683				
total_pemakaian	-0.047419				
rasio_pembayaran	-0.020056				
rasio_pembayaran_6bulan	-0.008569				
rasio_pembayaran_3bulan	-0.006857				
jumlah_tahun_sejak_pembukaan_kredit	-0.000222				
skor_delikuensi	0.003236				
total_pemakaian_tunai	0.016374				
pemakaian_6bln_per_limit	0.018036				
jumlah_kartu	0.023667				
limit_kredit	0.030994				
outstanding	0.089344				
utilisasi_6bulan	0.096429				
utilisasi_3bulan	0.103251				
tagihan	0.107561				
sisa_tagihan_per_jumlah_kartu	0.122347				
sisa_tagihan_tidak_terbayar	0.129020				
sisa_tagihan_per_limit	0.138163				
persentasi_overlimit	0.143785				
flag_kredit_macet	1.000000				

Feature Engineering

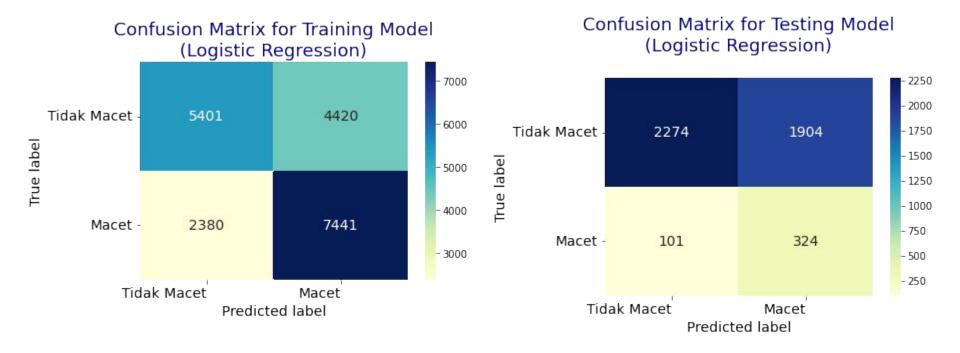
Feature Selection

jumlah kartu-	1	0.37	0.44	0.35	0.059	0.21	0.3	00025	0.047	0.012	0.012	0 014	0.024	0.39	0.21	011	0.03	0086	0065	0.013	0.046	0.02		1.0
jaman_karta	0.37	1	0.71	0.87		0.54	0.81				0.069				0.55	0.71		0.1	0.15	0.15	0.3			
limit_kredit	0.44	0.71	1	0.62	0.11	0.41	0.51	0.013	0007	0.014	0.11-	0.007	0.031	0.35	0.41	0.42	-0.11	0.029	0.074	0.059	-0.12	-0.13		
	0.35	0.87	0.62	1	0.21	0.34	0.83	0.008	0.26	.0003	90.12	0.017	0.11	0.18	0.35	0.74	0.31	0.015	0.14	0.14	0.32	0.23		0.8
total_pemakaian_tunai	0.059	0.24	0.11	0.21	1	0.23	0.17	.0009	0 .036	.4e-0	50004	0.005	0.016	0.0044	0.3	0.14	0.064	0.15	0.044	0.03	0.075	0.05		0.0
	0.21	0.54	0.41	0.34	0.23	1	0.21	0.014	0.028	0.012	0.023	-0.01	-0.05	0.11	1	0.15	0.059	0.47	0.18	0.075	0.0055	0.046		
sisa_tagihan_tidak_terbayar	0.3	0.81	0.51	0.83	0.17	0.21	1	0.021	0.29	0.004	0.081	0.016	0.13	0.18	0.22	0.91	0.44	0.028	0.064	0.12	0.36	0.28		
	0.002	.3e-0	0.013	0.008	.0009	9 .014	0.021	1	0.017	0.16	0.075-	0.002	2-0.02	0.0054	0.013	0.023	0.048	0.015	0.04	0.017	-0.027	0.021	8-	0.6
persentasi_overlimit	-0.047	0.23	.0007	0.26	0.036	0.028	0.29	0.017	1 -	0.002	0.054	0.013	0.14	0.023	0.031	0.3	0.5	0.047	0.24	0.26	0.51	0.4		
	0.012		0.014			50.012	SHEET OF		0.0022	1	0.29	ACT :	S		0,000000	Distance.	0.022		0.000	Spirit State				
rasio_pembayaran_6bulan					10000	4.700	6000	EXPLES	0.054	100000			100000	0.0097	02200		0.0002	200	100000	. 6. 6.	0.0035			
	10000	600000	No.		Scotton	00000	Contract,	SCALUS CO.	4000	1000000	5.0072		(0.0000)	000000	2000	200000	0.026		2000	10000	green to	400	- 1	0.4
flag_kredit_macet	-	1	0.031	NORTH BEAT	0.016		0.13	-0.02		Section 2	9.0086		1 4	era erang	D .047	0.12	0.14		0.067			0.096		
	0.39	0.2	0.35		0.0044	- A	Seattle Services			NAME OF TAXABLE	0.0097			A	0.11	0.13	0.092				-0.11	100		
total_pemakaian	Name and Address of the Owner, where			0.35	0.3	1					0.022			and the same of	1	0.15	0.053	-	0.18		0.011			
	0.11	0.71		0.74	0.14	0.15	0.91	-0.023			0.071			0.13	0.15	1		0.032		0.12	0.39	ACRONSOS.	8-	0.2
sisa_tagihan_per_limit		0.23	-0.11	0.31	0.064	0.059		0.048		0.022	3			0.092		0.48	1	0.028	0.15	0.21	0.78	0.63		
	0.0086		0.029		0.15	0.47	0. 32		0.047		0.015			CONTRACTOR OF	Service Co.	-	0.028	1	0.46	0.24		0.029		
pemakaian_3bln_per_limit			0.074 0.059		0.044			0.04	0.24		0.016					0.067		0.46	1	0.63	0.38			0.0
utilisasi 3bulan	0.013	0.15	-0.059	0.14		0.075		0.017 -0.027	0.26	0.026	1.0035				0.076 0.011	0.12	0.21	0.24	0.63	0.43	0.43	0.44		0.0
utilisasi_sbulari	0.046	0.21	-0.12	0.32	SHEELS	0.046		0.027	0.4	0.0088	0.008	CALL OF	3.3	2 80000	0.011	0.33	0.63	0.07	0.38	0.43	0.85	1		
	OCCUPANT.			5000				STATE OF THE PARTY							West of the	CALCULATION OF THE PARTY.			20024			156		
	umlah kartu	outstanding	limit_kredit	tagihan	tuna	reta	baya	rasio_pembayaran	persentasi_overlimit	bular	bular	skor_delikuensi	flag_kredit_macet	kredi	btal_pemakaian	kartı	Ē	Ē	Ē,	E.	utilisasi_3bulan -	utilisasi_6bulan -		
	nlah	ıtsta	ш	Ē	aian	aian	ter	mpa	ove	an_3	an_6	delik	dit	aan	ema	nlah	ber	per_	per	ber .	asi_3	asi_6		
	Ē	Б	_		emak	mak	tida	o pe	entas	ayar	ayar	sko	g_kre	nbuk	otal_p	Ţ	gihar	kaia	la S	- 6blr	utilis	SIII S		
					total_pemakaian_tunai	total_pemakaian_retail	sisa tagihan tidak terbayar	ras	perse	asio_pembayaran_3bulan	rasio_pembayaran_6bulan		Па	k per	Ħ	sisa_tagihan_per_jumlah_kartu	sisa tagihan per limit	btal_pemakaian_per_limit	pemakaian_3bIn_per_limit	pemakaian 6bln per limit				
					Ē	ā	a_tag			asio	asio			seja		agih	. CX	otal	sema	ema				
							Sis			_	Ŀ			ahun		sisa		-		_				
														jumlah tahun sejak pembukaan kredit										
														ш										

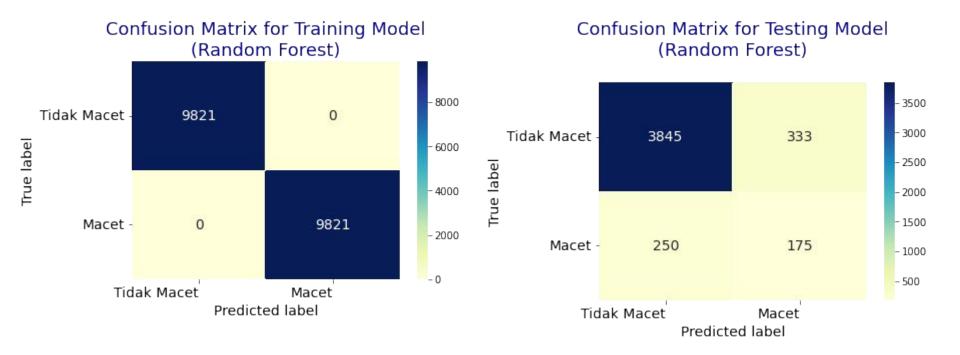


Korelasi Kuat

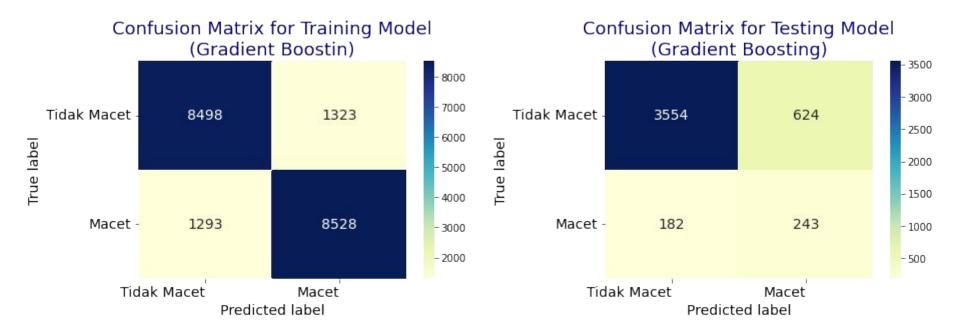
Beberapa variabel memiliki korelasi kuat (>0.7) satu sama lain sehingga dapat dihilangkan dan memilih beberapa variabel saja untuk dilakukan modeling.



- Data train: akurasi 65% dan recall 76%
- Data testing: akurasi 56% dan recall 76%
- Tidak terjadi peningkatan model score ketika dilakukan cross validation dan tuning parameter.



- Data train: akurasi dan recall 100%
- Data testing: akurasi 87% dan recall 41%



- Data train: akurasi dan recall 87%
- Data testing: akurasi 82% dan recall 52%

Algoritma	Akurasi Data Training	Akurasi Data Testing	Recall Data Training	Recall Data Testing	Performa
Logistic Regression	65%	56%	76%	76%	Kurang baik memprediksi di fase training dan testing
Random Forest Classifier	100%	87%	56%	56%	Over-Fitting
Gradient Boosting Classifier	87%	82%	87%	52%	Appropriate-Fitting

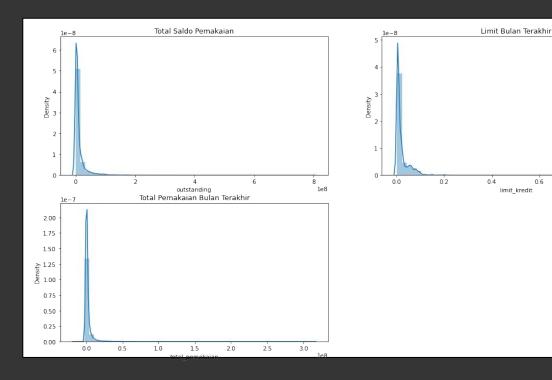
Load and Predict New Data

- Jumlah kartu terbanyak adalah 16 kartu.
- Dari beberapa kolom seperti total_pemakaian, rasio_pembayaran, dan total_pemakaian_per_limit, nilai minimum bernilai negatif.

0.6

0.8

1.0





Outlier

Terdapat perbedaan yang signifikan antara mean dan median dari kolom outstanding, limit_kredit, dan total_pemakaian yang mengindikasikan outlier atau skewed distribution.



Conclusion

- Untuk memprediksi kredit macet dari nasabah dengan menggunakan dataset ini model terbaiknya adalah menggunakan algortima Gradient Boosting Classifier.
- Hal ini dikarenakan performa dari model Gradient Boosting Classifier cenderung mampu memprediksi sama baiknya di fase training maupun testing (akurasi training 87%, akurasi testing 82%)
- Di lain sisi algoritma Logistic Regression kurang baik memprediksi di fase testing dan training serta algoritma Random Forest cenderung Under-Fitting.
- Akan tetapi hal ini tidak menjadikan kita untuk menarik kesimpulan bahwasannya jika untuk melakukan pemodelan apapun maka digunakan Gradient Boosting Classifier, kita tetap harus melakukan banyak percobaan model untuk menentukan mana yang terbaik.