



# LOAN APPROVAL PREDICTION

HENRY

PURWADHIKA STARTUP AND CODING SCHOOL  
DATA SCIENCE AND MACHINE LEARNING





# MACHINE LEARNING

Metode analisis data yang mengotomatiskan pembuatan model analitik, sebuah cabang kecerdasan buatan yang didasarkan pada gagasan bahwa sistem dapat belajar dari data, mengidentifikasi pola, dan membuat keputusan dengan intervensi manusia yang minimal.



# BANK LOAN

PENYEDIAAN UANG ATAU TAGIHAN-TAGIHAN BERDASARKAN PERJANJIAN PINJAM-MEMINJAM ANTARA BANK DENGAN PIHAK LAIN. PIHAK PEMINJAM BERKEWAJIBAN MELUNASI UTANGNYA SETELAH JANGKA WAKTU YANG TELAH DITETAPKAN DALAM PERJANJIAN.

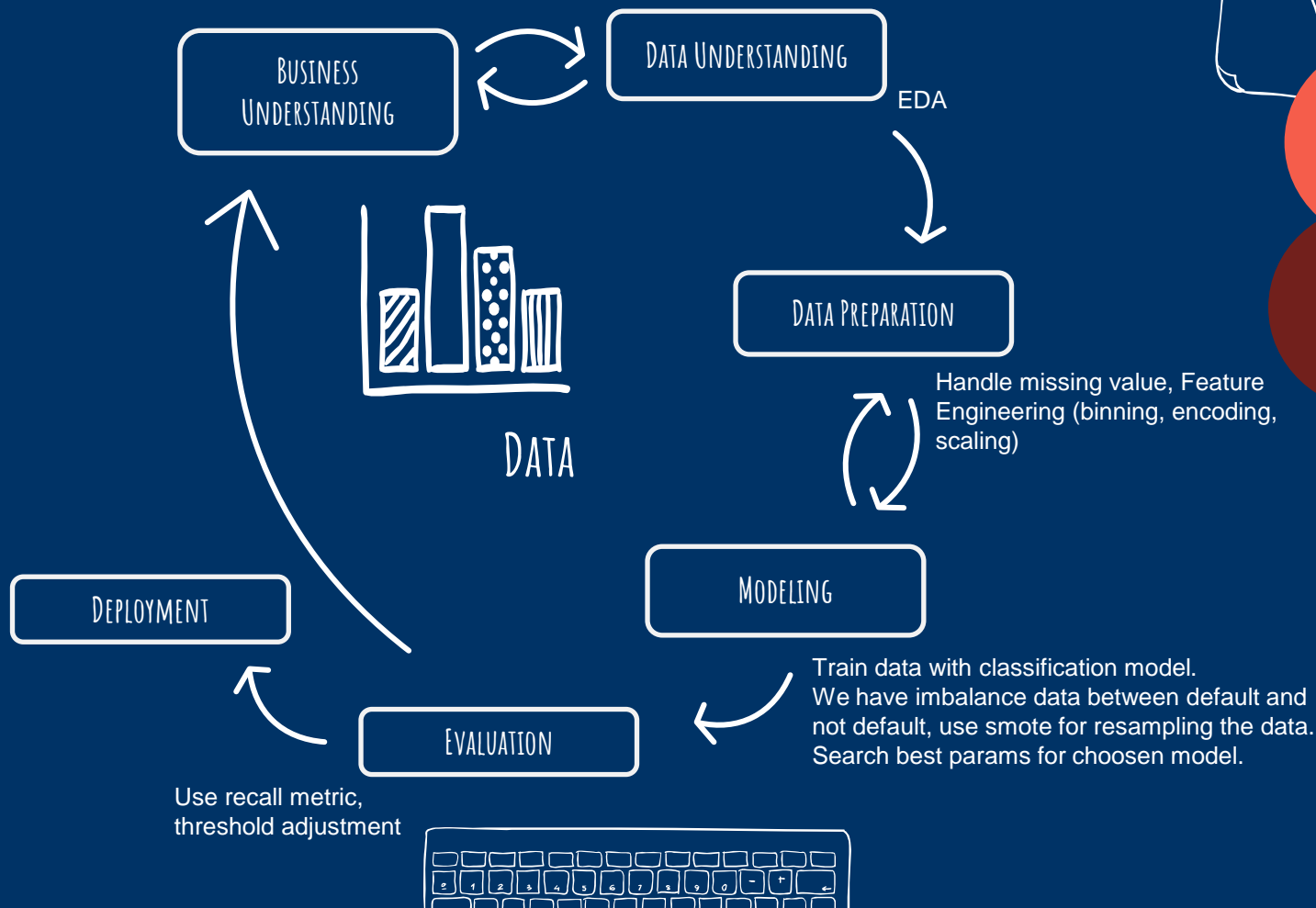


# BUSINESS PROBLEMS

Masih banyak Bank yang salah dalam memberikan pinjaman kepada nasabah. Banyak nasabah yang pada akhirnya gagal dalam melakukan pembayaran pinjaman. Hal tersebut diakibatkan oleh banyak faktor.

Tujuan dari machine learning ini adalah untuk meminimalisir resiko kesalahan pemberian pinjaman kepada nasabah.





# DATASET

Source : Kaggle ([Bank Loan Status Dataset](#) | Kaggle)

Shape of Dataset : (10514, 19)

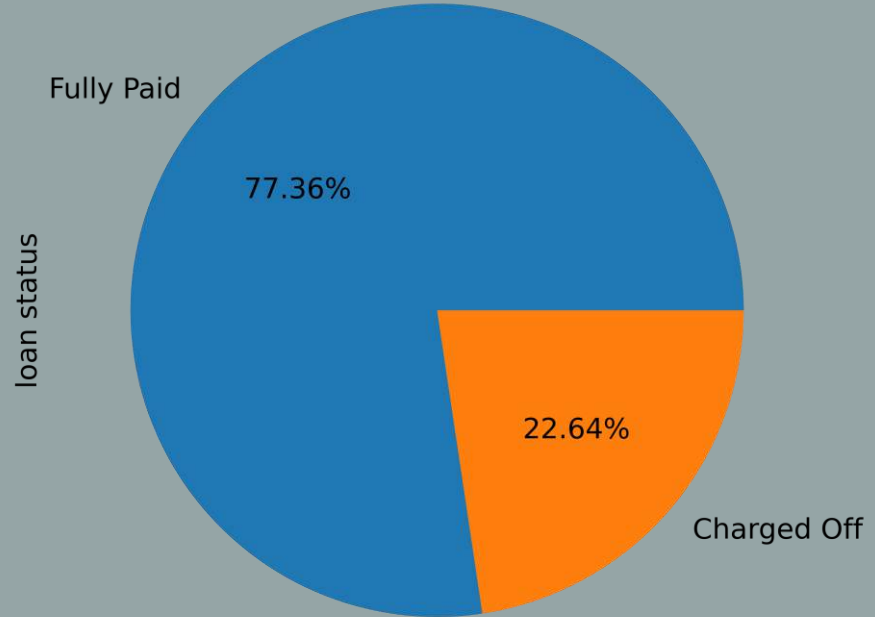
	loan id	customer id	loan status	current loan amount	term	credit score	annual income	years in current job	home ownership	purpose	monthly debt	years of credit history	months since last delinquent	number of open accounts	number of credit problems	current credit balance	maximum open credit	bankruptcies	tax liens
0	14cd8831-6af5-400b-83ec-68e61888a048	981165ec-3274-42f5-a3b4-d104041a9ca9	Fully Paid	445412.0	Short Term	709.0	1167493.0	8 years	Home Mortgage	Home Improvements	5214.74	17.2	NaN	6.0	1.0	228190.0	416746.0	1.0	0.0
1	4771cc26-131a-45db-b5aa-537ea4ba5342	2de017a3-2e01-49cb-a581-08169e83be29	Fully Paid	262328.0	Short Term	NaN	NaN	10+ years	Home Mortgage	Debt Consolidation	33295.98	21.1	8.0	35.0	0.0	229976.0	850784.0	0.0	0.0
2	4eed4e6a-aa2f-4c91-8651-ce984ee8fb26	5efb2b2b-bf11-4dfd-a572-3761a2694725	Fully Paid	99999999.0	Short Term	741.0	2231892.0	8 years	Own Home	Debt Consolidation	29200.53	14.9	29.0	18.0	1.0	297996.0	750090.0	0.0	0.0
3	77598f7b-32e7-4e3b-a6e5-06ba0d98fe8a	e777faab-98ae-45af-9a86-7ce5b33b1011	Fully Paid	347666.0	Long Term	721.0	806949.0	3 years	Own Home	Debt Consolidation	8741.90	12.0	NaN	9.0	0.0	256329.0	386958.0	0.0	0.0
4	d4062e70-befa-4995-8643-a0de73938182	81536ad9-5ccf-4eb8-befb-47a4d608658e	Fully Paid	176220.0	Short Term	NaN	NaN	5 years	Rent	Debt Consolidation	20639.70	6.1	NaN	15.0	0.0	253460.0	427174.0	0.0	0.0

# EDA

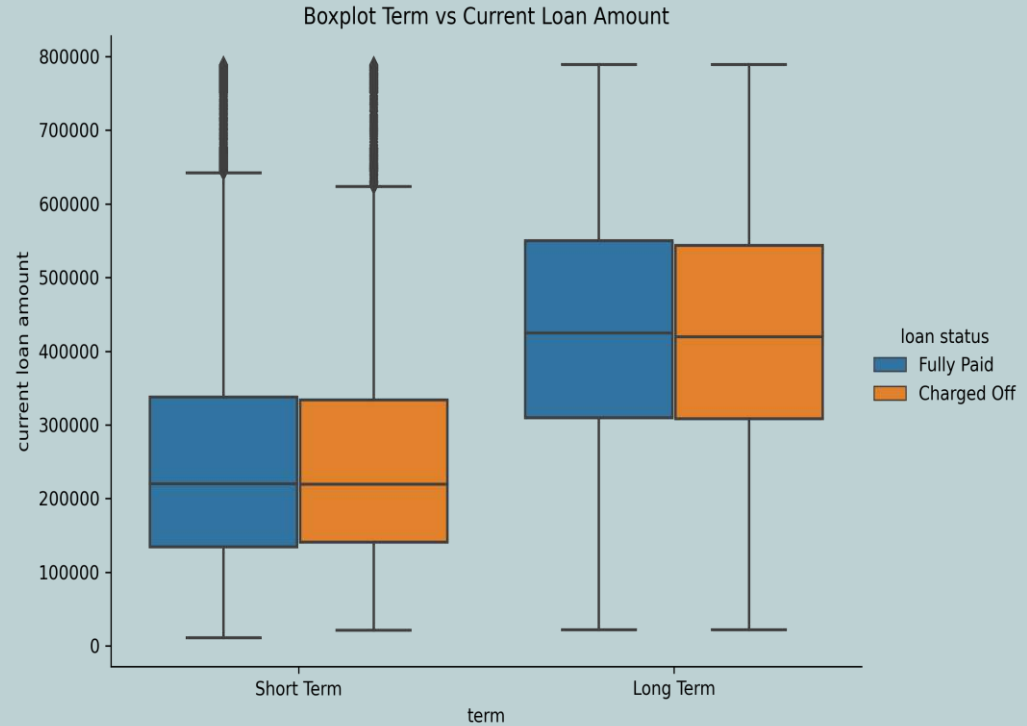
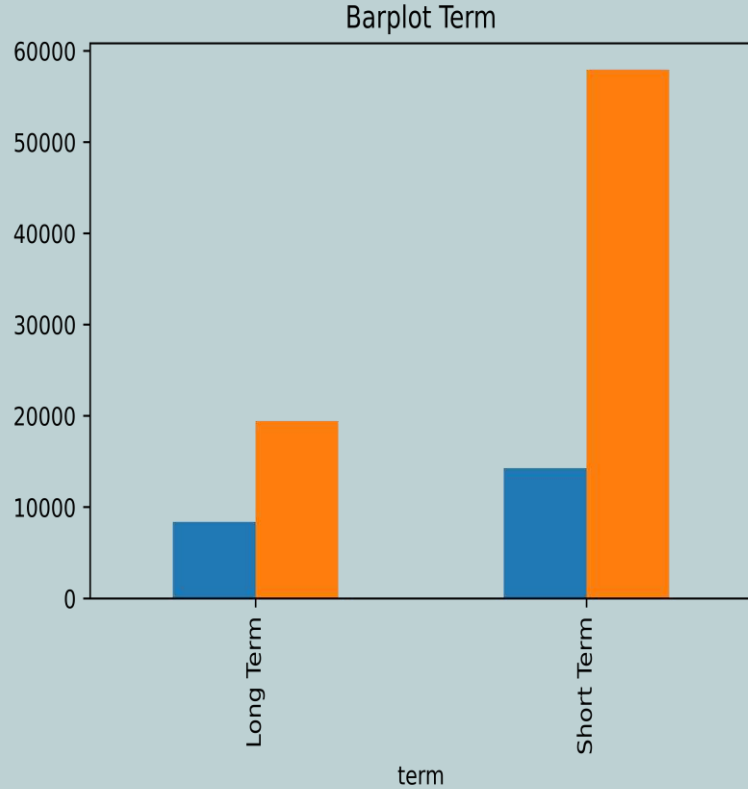
## PIEPLLOT LOAN STATUS

Persentase nasabah yang berhasil bayar sebanyak 77.36% dibandingkan nasabah yang gagal bayar

Piechart Loan Status



# GRAFIK LOAN STATUS BERDASARKAN TERM

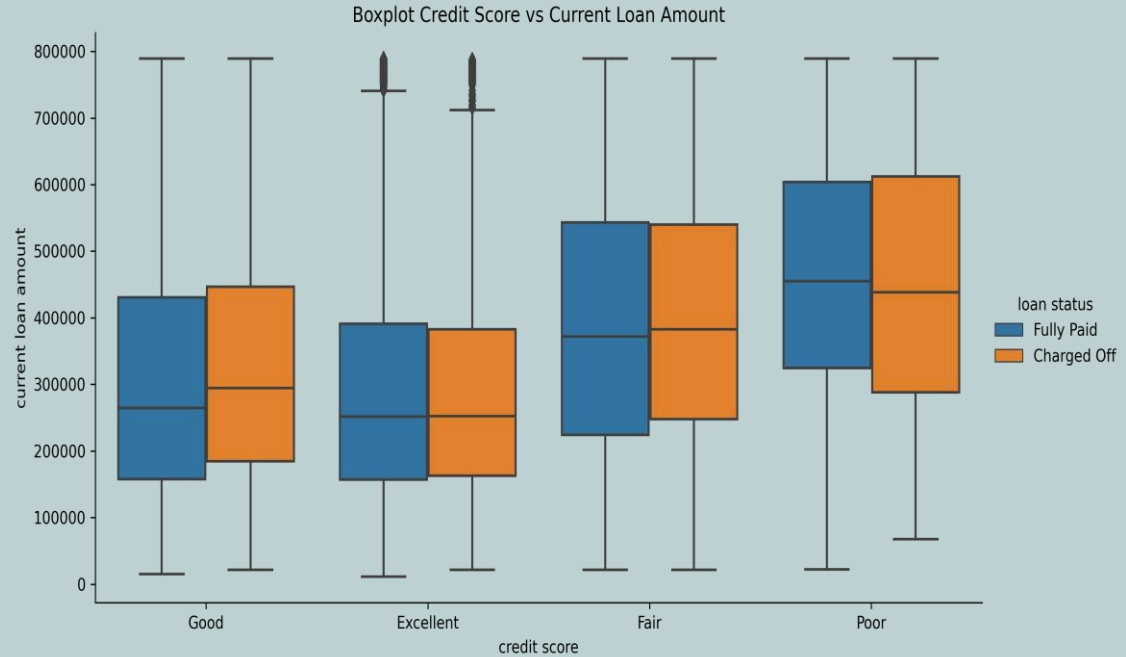
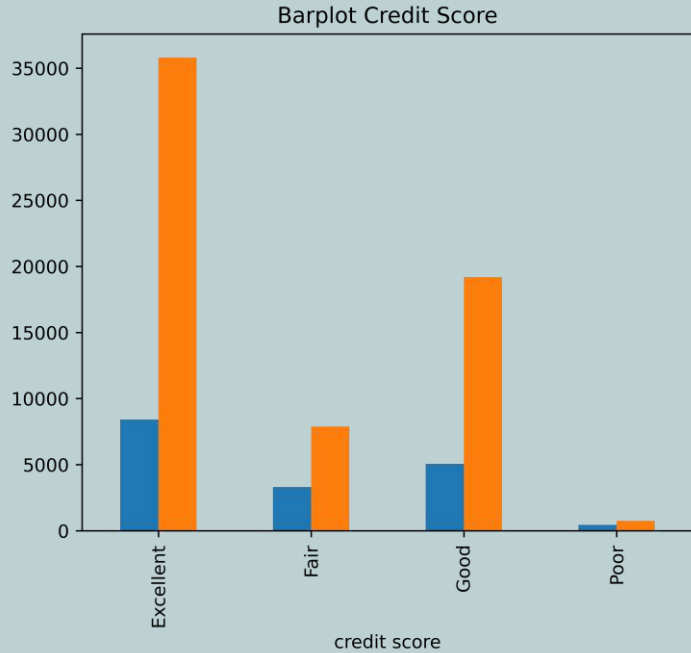


Persentase orang yang gagal bayar pada long term lebih tinggi dibandingkan pada short term



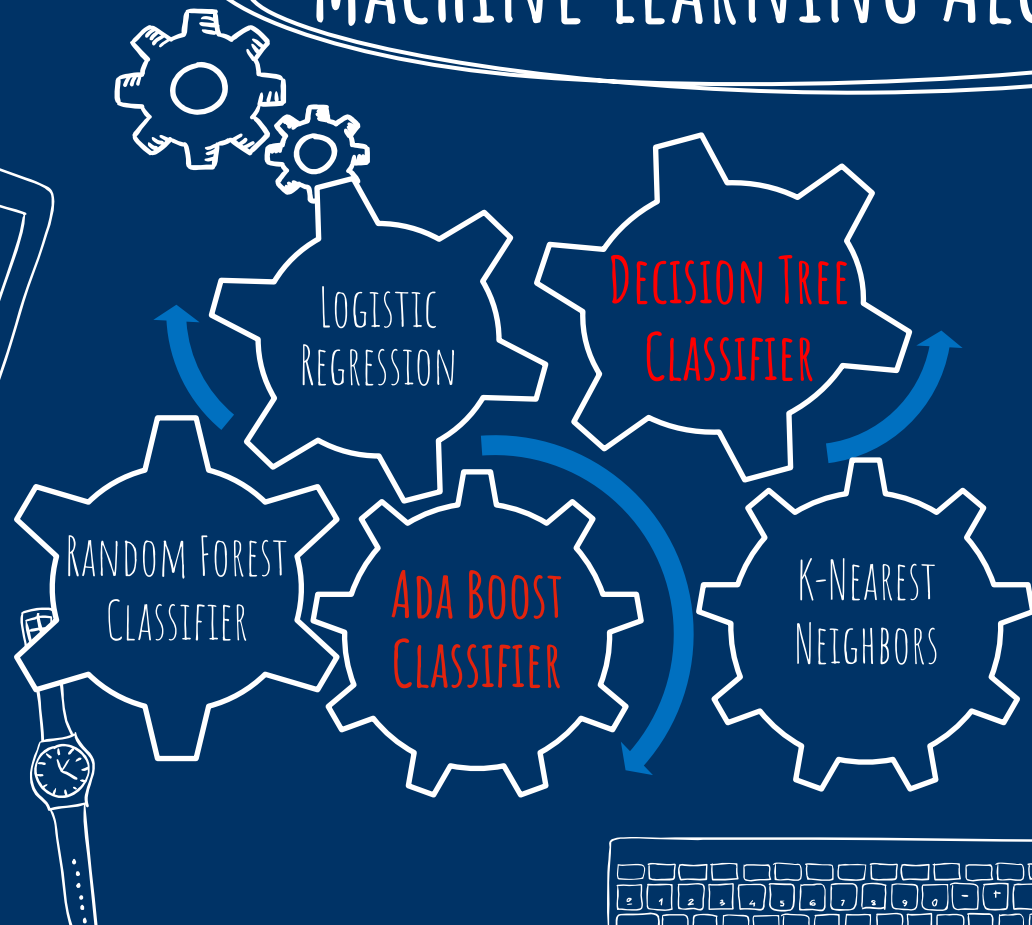
# GRAFIK LOAN STATUS BERDASARKAN CREDIT SCORE

9



Semakin buruk credit score nasabah, semakin tinggi persentase nasabah tersebut gagal bayar

# MACHINE LEARNING ALGORITHM



Recall Score tiap model :

Logistic Regression Score : 0.5762

Random Forest Classifier Score : 0.6448

**Decision Tree Classifier Score : 0.6680**

K-Nearest Neighbors Score : 0.4754

**AdaBoost Classifier Score : 0.6944**

# RECALL & PRECISION

## RECALL

**Recall** dapat didefinisikan sebagai rasio dari *jumlah total contoh positif yang diklasifikasikan bernilai benar* dibagi dengan *jumlah total contoh positif*. **High Recall** menunjukkan kelas dikenali dengan baik (FN rendah).

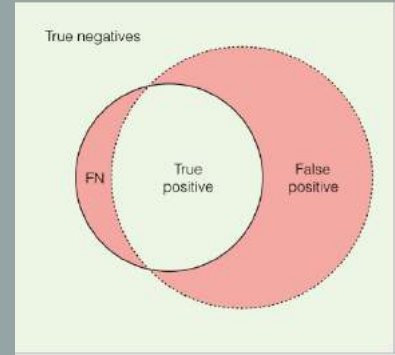
$$Recall = \frac{TP}{TP + FN}$$

## PRECISION

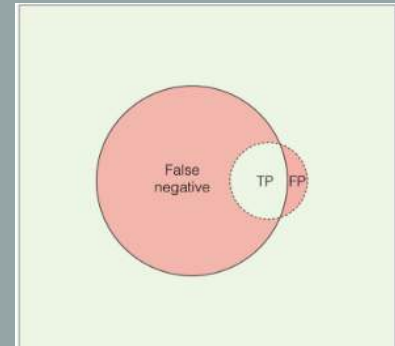
**Precision** merupakan pembagian dari *jumlah total contoh positif yang diklasifikasikan bernilai benar* dengan *jumlah total contoh positif yang diprediksi*. **High Precision** menunjukkan contoh berlabel positif memang positif (FP rendah).

$$Precision = \frac{TP}{TP + FP}$$

HIGH RECALL, LOW PRECISION

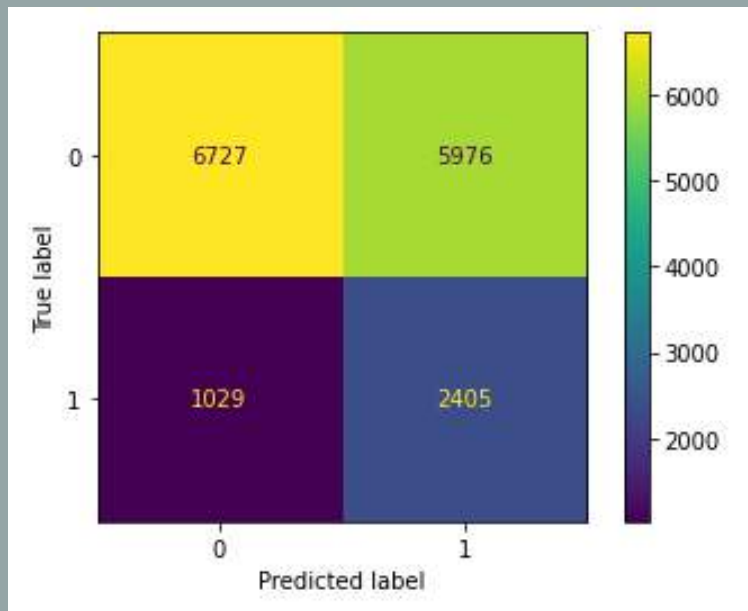


LOW RECALL, HIGH PRECISION:

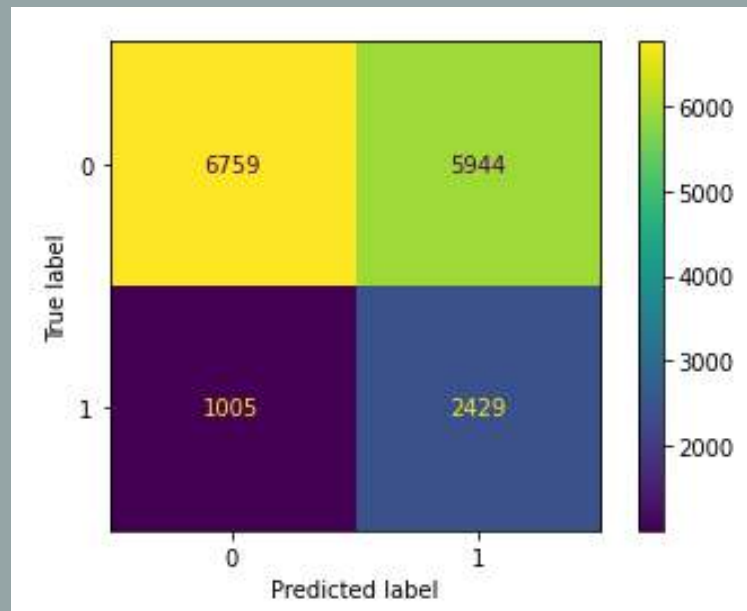


# CONFUSION MATRIX BETWEEN DTC DAN ABC

Decision Tree Classifier

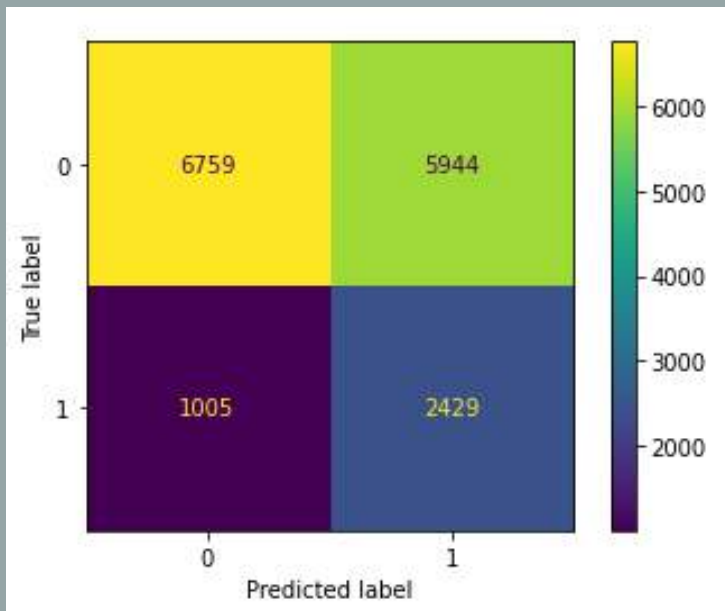


Ada Boost Classifier

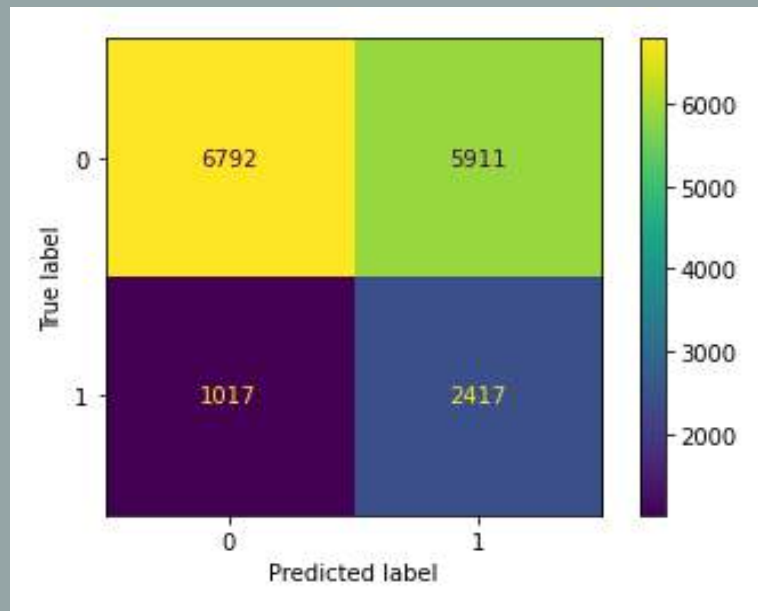


# CONFUSION MATRIX ABC AFTER HYPERPARAMETER TUNING

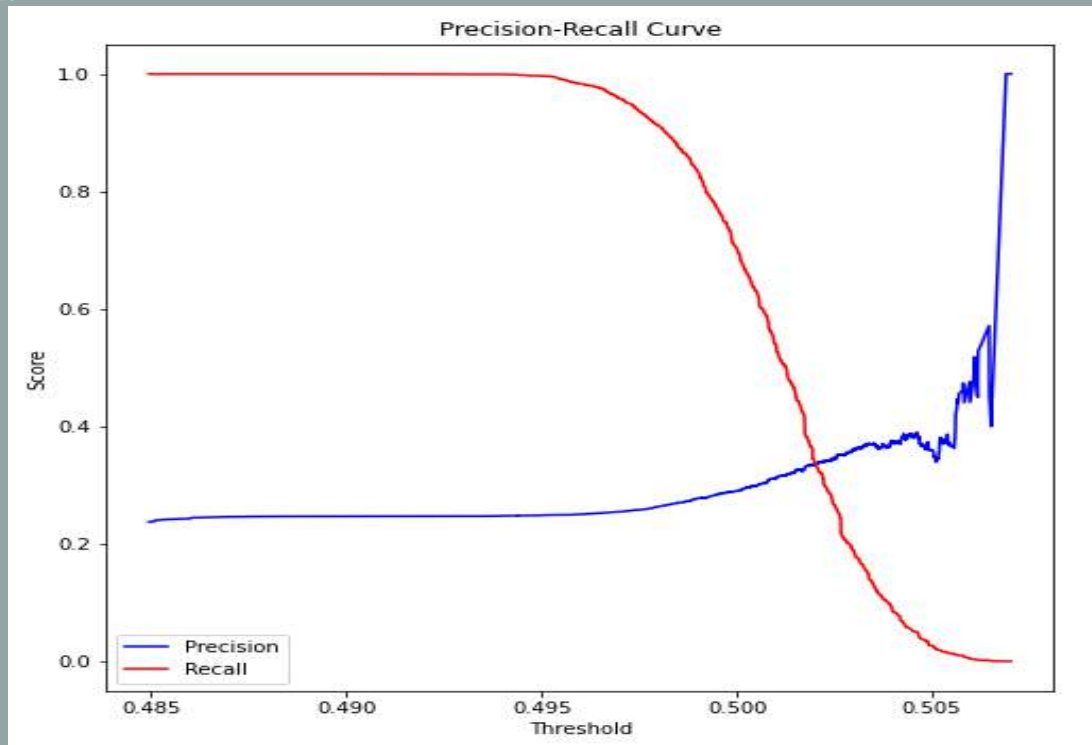
Before Hyperparameter Tuning



After Hyperparameter Tuning



# THRESHOLD ADJUSTMENT BERDASARKAN PRECISION-RECALL CURVE



Threshold yang digunakan bernilai 0.4990552628825836, hal ini berarti probability di bawah nilai tersebut akan diklasifikasikan sebagai 0 dan probability di atas nilai tersebut akan diklasifikasikan sebagai 1.

# CONFUSION MATRIX AFTER THRESHOLD ADJUSTMENT

## Confusion Matrix

	prediction		total actual
	0	1	
0	6759	5944	12703
1	1005	2429	3434
total prediction	7764	8373	16137

## Confusion Matrix

	prediction		total actual
	0	1	
0	5339	7364	12703
1	592	2842	3434
total prediction	5931	10206	16137

# CHOOSEN ALGORITHM

## ADA BOOST CLASSIFIER

### Confusion Matrix

	prediction		total actual
	0	1	
0	5339	7364	12703
1	592	2842	3434
total prediction	5931	10206	16137

### Classification Report

	precision	recall	f1-score	support
0	0.89	0.44	0.59	12703
1	0.28	0.81	0.41	3434
accuracy			0.52	16137
macro avg	0.59	0.62	0.50	16137
weighted avg	0.76	0.52	0.55	16137





## CONCLUSION

### Modeling Report

Ada Boost Classifier merupakan model yang akhirnya dipakai. Model tersebut memiliki nilai recall 0.81 dan precision 0.28

Dengan mengatur threshold, kita dapat mengurangi potensi untuk memberikan pinjaman kepada nasabah yang salah sebesar 2.96%

