

**BREAK THE
LIMIT !!!**



Bank Marketing Campaign

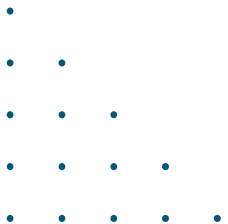
CAPSTONE PROJECT
PURWADHIKA

BY. Ahnaf Indrastata

LATAR BELAKANG

Bank merupakan sebuah lembaga intermediasi keuangan umumnya didirikan dengan kewenangan menerima simpanan uang, meminjamkan uang, dan menerbitkan promes atau yang dikenal sebagai banknote. Kata bank berasal dari bahasa Italia banca yang berarti tempat penukaran uang.

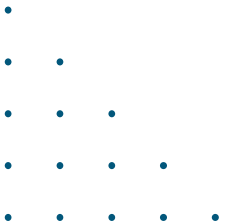
Di era yang berkembang seperti saat ini, Bank harus bisa bersaing dengan kompetitor-kompetitor lainnya. Salah satu caranya dengan strategi pemasaran.



Problem Statement

Dalam menaikkan profit bank, maka bank memerlukan strategy dalam meningkatkan profit. Salah satunya dengan melakukan kampanye pemasaran. Saat ingin melakukan kampanye pemasaran bank memerlukan targa kepada semua nasabahnya. Cara itu membuat expense perusahaan melonjak. Oleh karena itu saat dilakukan pemasaran, bank harus bisa melakukan target yang tepat sasaran.

Untuk meningkatkan ketepatan dan efesiensi yang maksimal dalam kampanye pemasaran, perusahaan juga harus mencari peluang target yang tinggi.



Goals

Banyaknya nasabah yang dimiliki bank sangat menghabiskan waktu dan memerlukan biaya tambahan jika perusahaan menggunakan vendor atau jasa dalam menyaring nasabah yang memiliki tepat sasaran. Oleh sebab itu perusahaan bank harus bisa mem-forecasting nasabah dengan karakter seperti apa yang bisa menjadi target dalam melakukan kampanye pemasaran.

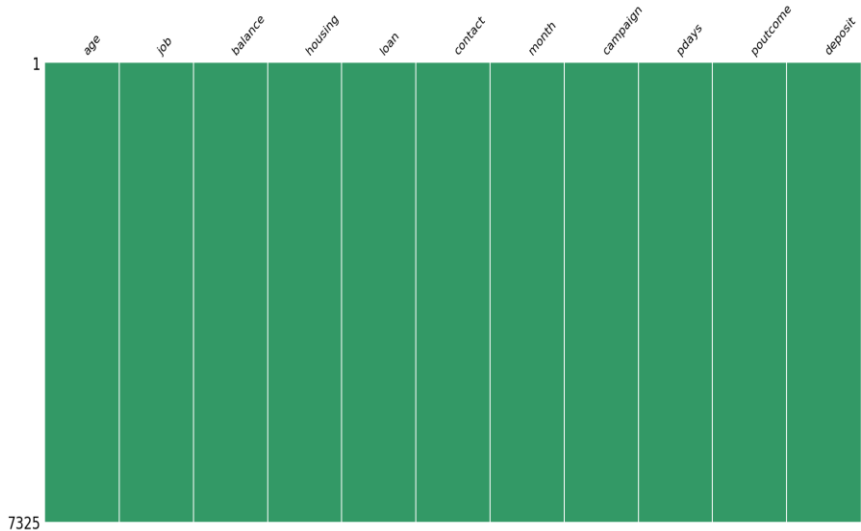
Setelah itu perusahaan dapat mengetahui profile atau segmen mana saja yang membuat nasabah ingin atau tidak melakukan deposit. Sehingga perusahaan bisa membuat semacam Business Plan berdasarkan pengelompokan target nasabah yang sudah dibuat.

- •
- • •
- • • •
- • • • •

BREAK THE
LIMIT !!!

DATA SET

7813 rows × 11 columns



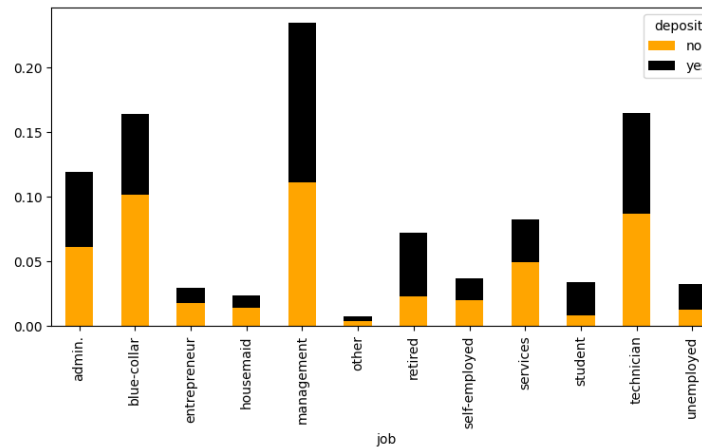
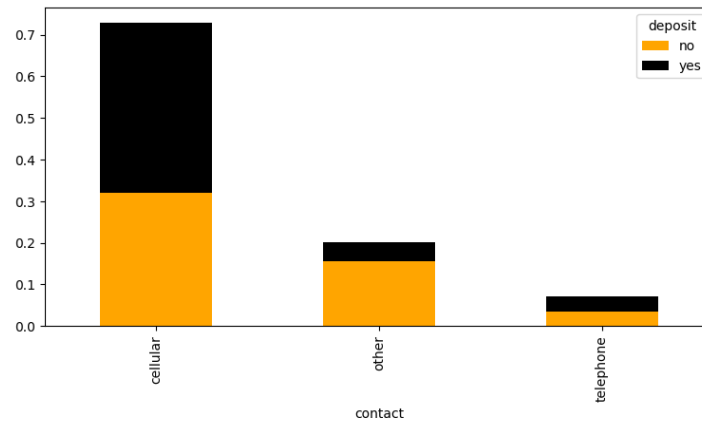
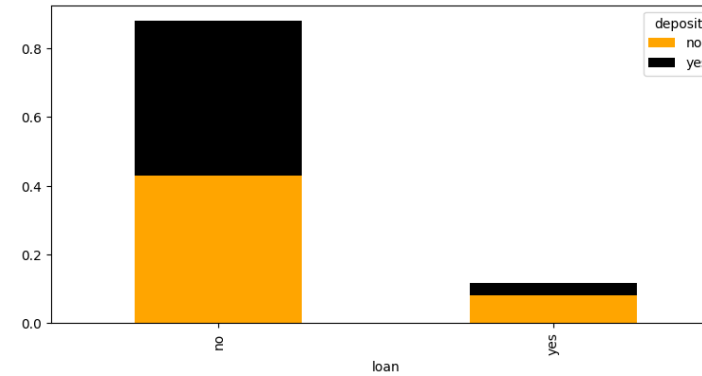
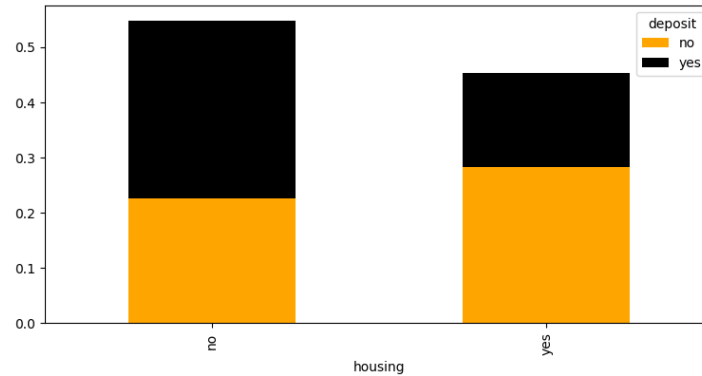
'Total yang mempunyai data duplikat : 8'

-
- •
- • •
- • • •
- • • • •

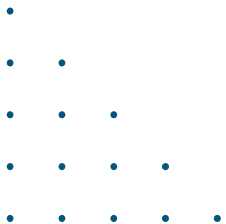
<i>Attribute</i>	<i>Data Type</i>	<i>Description</i>
age	Integer	Usia
job	Object	Pekerjaan
balance	Integer	Saldo
housing	Object	Mempunyai kredit Rumah atau tidak
loan	Object	Mempunyai pinjaman atau tidak
contact	Object	Tindakan komunikasi terbaru yang dilakukan terhadap nasabah.
month	Object	Bulan terakhir berhubungan dengan nasabah.
campaign	Integer	Total interaksi yang dilakukan dengan nasabah selama kampanye.e
pdays	Integer	Total hari setelah nasabah dihubungi sejak kampanye terakhir
poutcome	Object	Hasil dari kampanye terakhir
deposit	Object	Melakukan deposit atau tidak

BREAK THE
LIMIT !!!

Data Analysis



*Pengaruh terhadap deposit (Yes/No)



Data Preprocessing



	count	mean	std	min	25%	50%	75%	max
age	7325.0	41.356860	12.052640	18.0	32.0	39.0	49.0	95.0
balance	7325.0	1626.027577	3057.490750	0.0	187.0	620.0	1808.0	52587.0
campaign	7325.0	2.506348	2.713091	1.0	1.0	2.0	3.0	63.0
pdays	7325.0	52.396451	108.360776	-1.0	-1.0	-1.0	77.0	854.0
deposit	7325.0	0.490102	0.499936	0.0	0.0	0.0	1.0	1.0

Berdasarkan value masing-masing feature baik itu yang numerikal maupun kategorikal, kita akan melakukan :

- Feature housing, loan, contact dan poutcome akan kita lakukan One Hot Encoding
- Feature month dilakukan Ordinal Encoding.
- Feature job dilakukan Binary Encoding.
- Feature age, balance, campaign dan pdays dilakukan Robust Scaling.

	Column Name	Number of Unique	Unique Sample
0	job	12	[admin., housemaid, technician, management, student, services, blue-collar, entrepreneur, retired, unemployed, self-employed, other]
1	housing	2	[no, yes]
2	loan	2	[no, yes]
3	contact	3	[cellular, telephone, other]
4	month	12	[jun, may, nov, jan, sep, feb, mar, aug, apr, jul, oct, dec]
5	poutcome	4	[unknown, other, failure, success]

Benchmarking

- Ada Boost memiliki nilai paling tinggi pad F2-Score
- Hasil ini memberikan gambaran tentang kinerja model pada data pelatihan dan validasi. Rata-rata dari masing-masing metrik evaluasi juga disediakan, memberikan indikasi tentang konsistensi kinerja model secara keseluruhan selama proses cross-validation.
- Dengan demikian, Anda dapat memeriksa apakah model memiliki overfitting atau underfitting, serta membandingkan performa model dengan berbagai metrik evaluasi yang relevan.

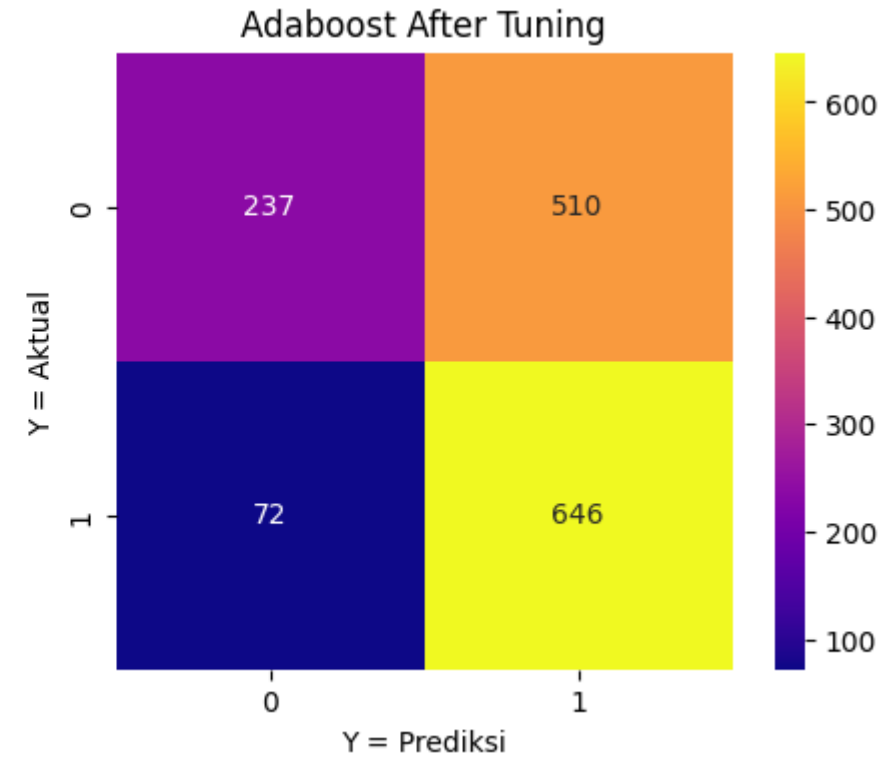
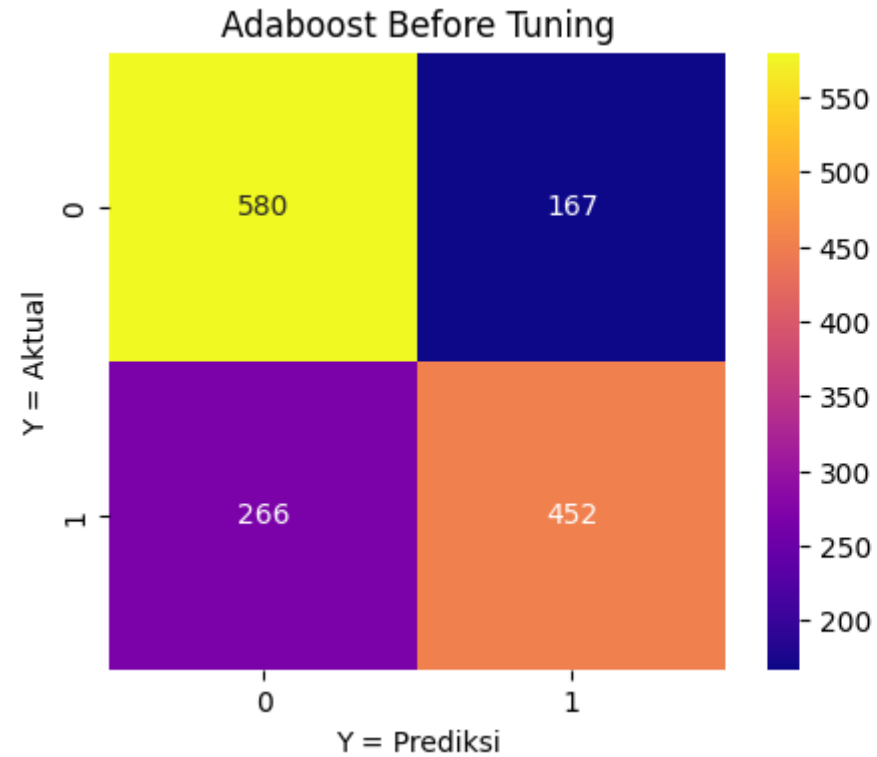
	models	fit_time	score_time	test_accuracy	test_precision	test_recall	test_f1	test_f2
5	Ada Boost	0.271772	0.036259	0.710068	0.728800	0.651469	0.687849	0.665525
0	Logistic Regression	0.095795	0.035973	0.677816	0.674546	0.661900	0.668044	0.664317
4	Cat Boost	3.501971	0.035905	0.718771	0.746408	0.645532	0.692190	0.663395
3	Random Forest	0.869491	0.067536	0.695563	0.705033	0.651461	0.677142	0.661486
7	Gradient Boosting	0.650758	0.026075	0.723208	0.758739	0.638571	0.693307	0.659360
6	LGBM Classifier	0.151031	0.032732	0.709044	0.738604	0.629171	0.679424	0.648337
8	XGBoost	0.268253	0.038490	0.690785	0.705079	0.634741	0.667986	0.647618
1	K-Nearest Neighbor	0.048981	0.115961	0.659556	0.662388	0.622898	0.641957	0.630367
2	Decision Tree Classifier	0.069452	0.024972	0.625085	0.618335	0.617683	0.617737	0.617642

-
- •
- • •
- • • •
- • • • •

	Train Accuracy	Test Accuracy	Train Precision	Test Precision	Train Recall	Test Recall	Train F1	Test F1	Train F2	Test F2
0	0.716297	0.708191	0.732916	0.722222	0.662750	0.656794	0.696069	0.687956	0.675688	0.668914
1	0.719710	0.726109	0.740469	0.748527	0.659269	0.663763	0.697514	0.703601	0.674052	0.679144
2	0.715444	0.704778	0.733075	0.718391	0.659704	0.653310	0.694457	0.684307	0.673179	0.665366
3	0.720350	0.724403	0.755440	0.757143	0.634741	0.645217	0.689851	0.696714	0.655693	0.664875
4	0.727176	0.686860	0.753234	0.697719	0.659121	0.638261	0.703042	0.666667	0.676014	0.649328
Average	0.719795	0.710068	0.743027	0.728800	0.655117	0.651469	0.696187	0.687849	0.670925	0.665525

**BREAK THE
LIMIT !!!**

Before - After



Conclusion

Berdasarkan analisis kesimpulan yang dapat diambil:

Performa Keseluruhan Model:

- Dengan tingkat akurasi sebesar 60,3%, dapat disimpulkan bahwa model yang dibuat masih belum cukup baik dalam memprediksi kedua kelas, yaitu nasabah yang melakukan deposit dan yang tidak. Hal ini disebabkan oleh adanya trade-off antara recall dan presisi. Meskipun fokus utama adalah meningkatkan prediksi yang benar untuk nasabah yang akan deposit, namun prediksi yang benar untuk nasabah yang tidak deposit menjadi berkurang. Error rate pada model tersebut mencapai 39,7%.
- Dengan F2-Score sekitar 80,2%, model terbilang cukup baik dalam memprediksi nasabah yang akan melakukan deposit. F2-Score memberikan bobot yang lebih besar pada recall, sehingga lebih memprioritaskan prediksi yang benar untuk kelas positif.

Feature Penting dalam Klasifikasi:

- Feature yang paling penting dalam klasifikasi nasabah yang akan deposit dan tidak deposit adalah `contact_other`, yang menunjukkan jenis komunikasi terakhir selain `cellular` dan `telephone`. Selanjutnya, `feature_poutcome_success`, yang menandakan hasil sukses dari kampanye pemasaran terakhir, juga memiliki kontribusi yang signifikan dalam klasifikasi.

1156 nasabah hanya 646 melakukan deposit:

Total waktu yang digunakan: 1156×10 (menit) = 192jam

- Total biaya yang dikeluarkan: $1156 \times \$5 = \5780
- Kerugian: $(1156 - 646) \times \$5 = \2.550



Recommendation

Dalam mengembangkan proyek ini lebih lanjut dan meningkatkan kualitas modelnya, beberapa langkah yang dapat diambil adalah:

Menggunakan Hyperparameter yang lebih baik:

- Melakukan pengoptimalan lebih lanjut terhadap hyperparameter model, baik melalui metode grid search atau teknik optimasi lainnya, untuk memastikan bahwa model mencapai performa optimalnya.

Penambahan Fitur Baru:

- Menambahkan fitur-fitur baru seperti kisaran gaji, status perkawinan, dan jumlah anak dapat memberikan informasi tambahan yang relevan dalam memprediksi minat calon nasabah untuk melakukan deposit. Fitur-fitur ini dapat memberikan pemahaman yang lebih baik tentang kondisi keuangan dan tanggung jawab keuangan calon nasabah, sehingga memungkinkan model untuk membuat prediksi yang lebih akurat.

Analisis Lebih Lanjut terhadap Data:

- Melakukan analisis lebih lanjut terhadap data untuk memahami lebih dalam pola-pola dan insight-insight yang terkandung di dalamnya. Hal ini dapat dilakukan melalui eksplorasi data yang lebih mendalam, visualisasi data yang lebih kompleks, atau penggunaan teknik analisis statistik yang lebih canggih.

Melakukan analisa data dari model data yang kita kurang tepat saat melakukan forecasting data. Sehingga kita mampu mengetahui bagaimana karakteristik dan alasannya.