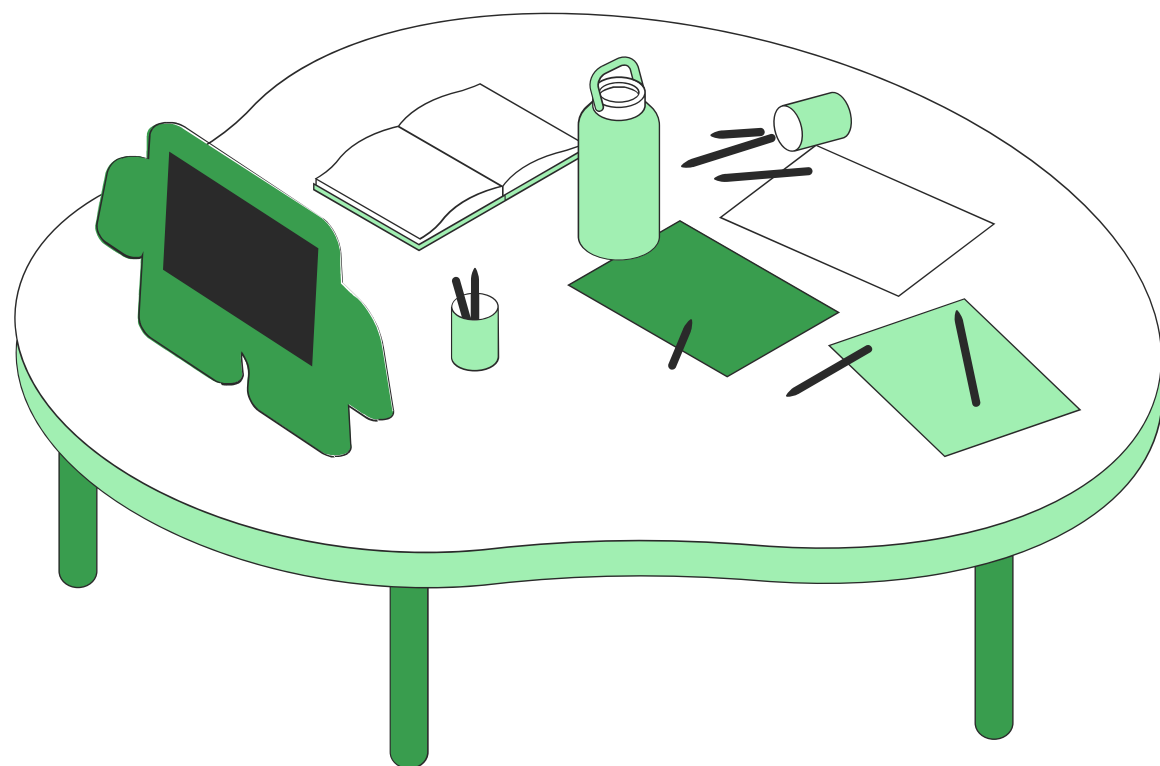


Prediksi Nasabah Potensial untuk Produk Deposito Berjangka

Adi Nugraha, JCDS Office Hour



Table of Content



Latar Belakang Masalah

Tujuan Proyek

Mengapa ML

Deskripsi Dataset

Temuan Kunci dari EDA

Persiapan Data

Pemilihan Fitur

Transformasi Fitur

Pendekatan Pemodelan

Hasil Evaluasi Model

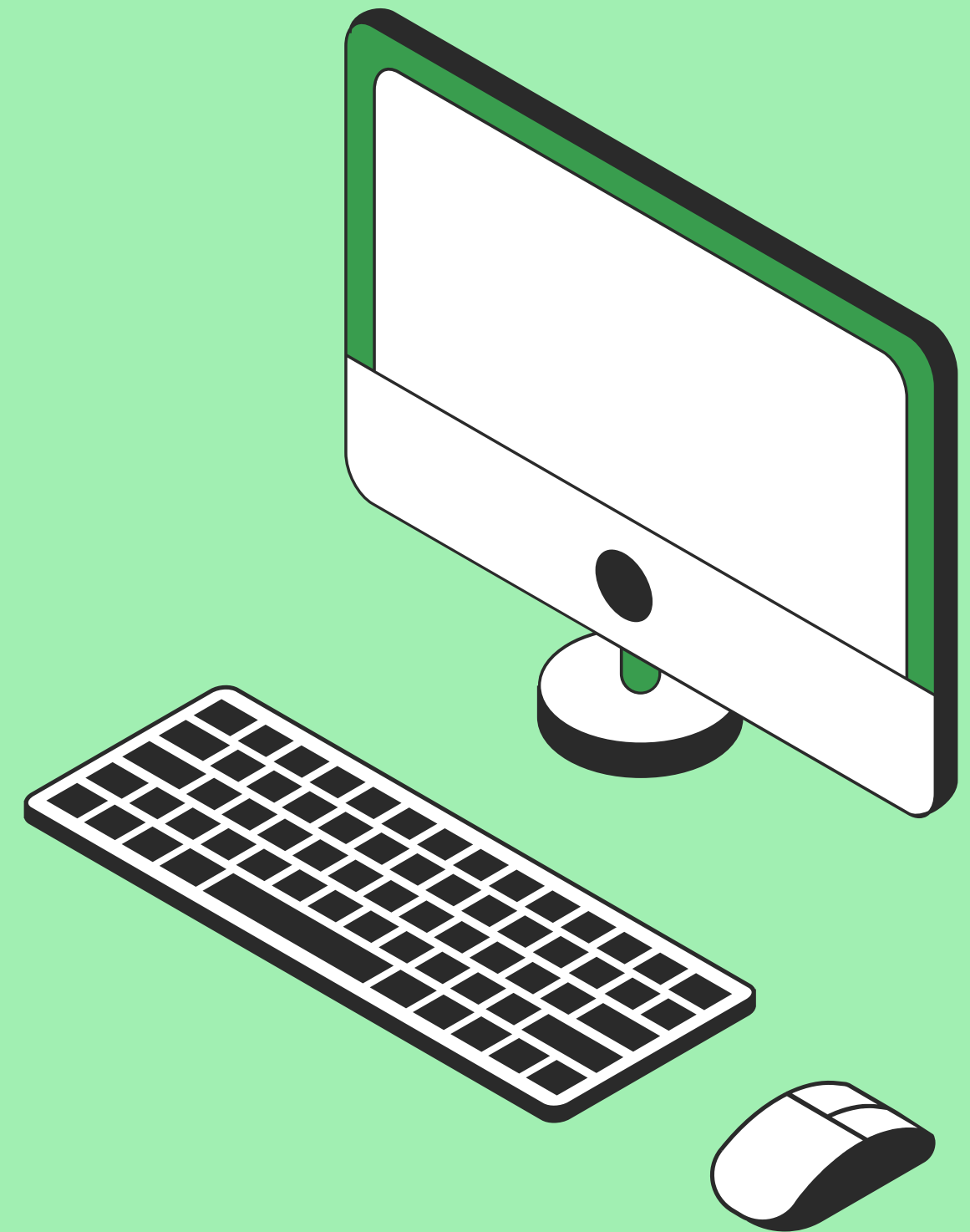
Kesimpulan

Rekomendasi Bisnis

Latar Belakang Masalah

Tantangan Kampanye Pemasaran

Saat ini, bank menghadapi tantangan dalam menjalankan kampanye pemasaran untuk produk **deposito**. Meskipun jangkauannya luas, pendekatan yang digunakan belum sepenuhnya tertarget, sehingga mengakibatkan **biaya operasional yang tinggi** dan **tingkat konversi yang belum optimal**. Banyak **waktu** dan **sumber daya terbuang** untuk menghubungi nasabah yang sebenarnya tidak memiliki potensi, yang pada akhirnya mengurangi efektivitas kampanye secara keseluruhan.



Apa yang Ingin Dicapai?

Tujuan Proyek



Goal Utama

Membangun model klasifikasi untuk memprediksi nasabah yang paling mungkin berlangganan deposito.



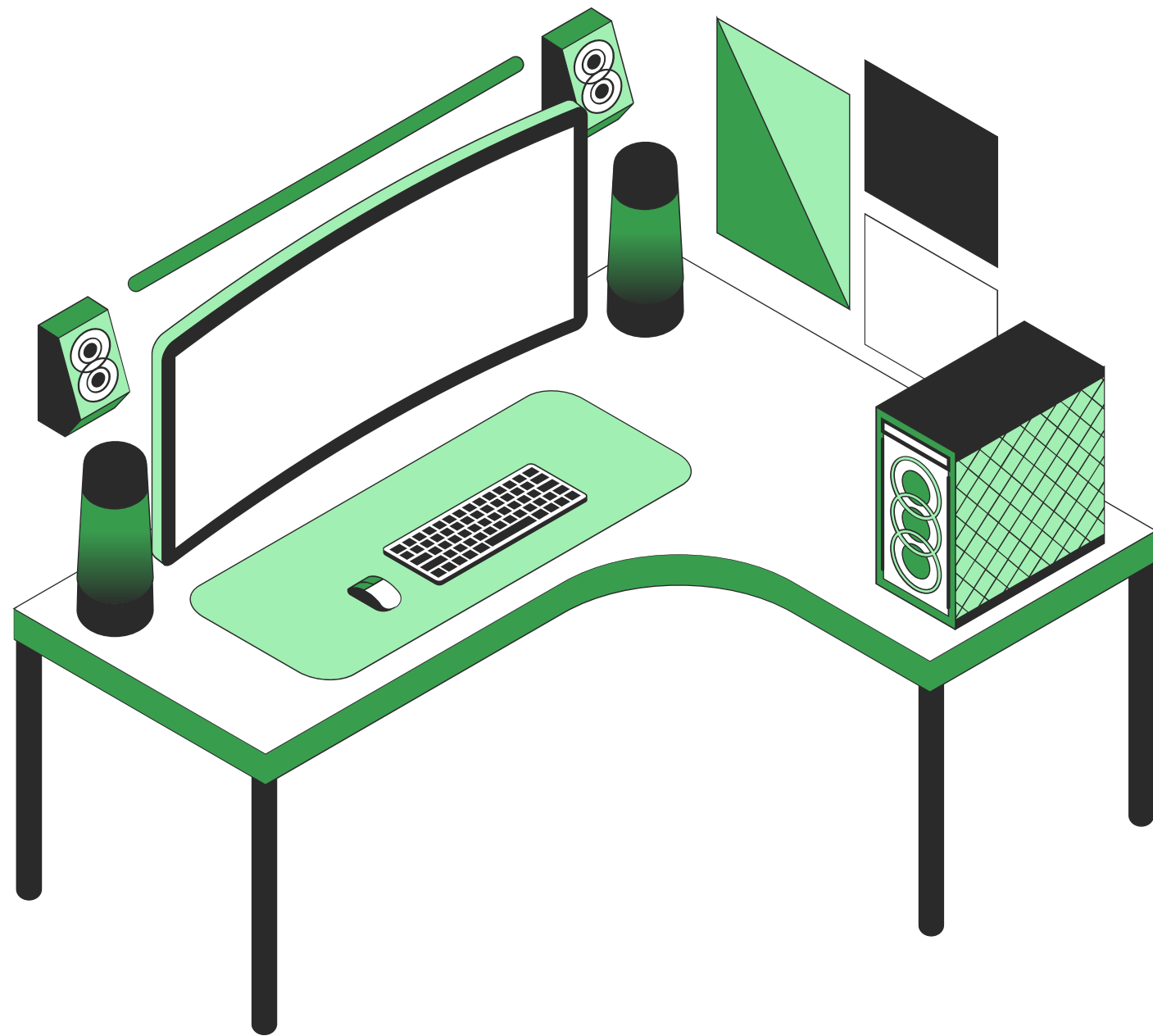
Goal Bisnis

Meningkatkan efisiensi kampanye dengan memfokuskan sumber daya pada target yang tepat.



Goal Analitis

Mengidentifikasi faktor-faktor kunci yang memengaruhi keputusan nasabah.



Bagaimana Caranya?

Mengapa ML

Untuk mencapai tujuan yang telah ditetapkan, pendekatan manual atau **aturan sederhana tidak akan cukup**. Analisis ini dihadapkan pada ribuan data nasabah dengan **kombinasi fitur yang sangat kompleks**. Oleh karena itu, **machine learning** diperlukan karena kemampuannya untuk **menganalisis data dalam skala besar** dan menemukan **pola-pola tersembunyi** yang tidak dapat dilihat manusia. Penerapan machine learning memungkinkan pembangunan sebuah sistem otomatis yang tidak hanya **memprediksi nasabah potensial** saat ini, tetapi juga dapat terus **beradaptasi untuk kampanye-kampanye di masa depan** secara efisien.

Memahami Data Nasabah

Deskripsi Dataset

Kolom	Penjelasan
age	Usia nasabah dalam tahun. Ini adalah fitur numerik.
job	Jenis pekerjaan nasabah (contoh: 'admin.', 'technician', 'services'). Ini adalah fitur kategorikal.
balance	Saldo rata-rata tahunan nasabah dalam mata uang Euro. Ini adalah fitur numerik.
housing	Menunjukkan apakah nasabah memiliki kredit perumahan ('yes' atau 'no'). Ini adalah fitur kategorikal.
loan	Menunjukkan apakah nasabah memiliki pinjaman pribadi ('yes' atau 'no'). Ini adalah fitur kategorikal.
contact	Jenis media komunikasi yang digunakan untuk menghubungi nasabah (contoh: 'cellular', 'telephone'). Ini adalah fitur kategorikal.
month	Bulan terakhir nasabah dihubungi dalam setahun (contoh: 'jan', 'feb', 'mar', dst.). Ini adalah fitur kategorikal.
campaign	Jumlah kontak yang dilakukan selama kampanye ini untuk nasabah tersebut. Ini adalah fitur numerik.
pdays	Jumlah hari yang berlalu setelah nasabah terakhir dihubungi dari kampanye sebelumnya (-1 berarti nasabah belum pernah dihubungi). Ini adalah fitur numerik.
poutcome	Hasil dari kampanye pemasaran sebelumnya (contoh: 'success', 'failure', 'other', 'unknown'). Ini adalah fitur kategorikal.
deposit	(TARGET) Variabel yang akan kita prediksi. Menunjukkan apakah nasabah akhirnya berlangganan deposito berjangka ('yes') atau tidak ('no'). Ini adalah target klasifikasi kita.

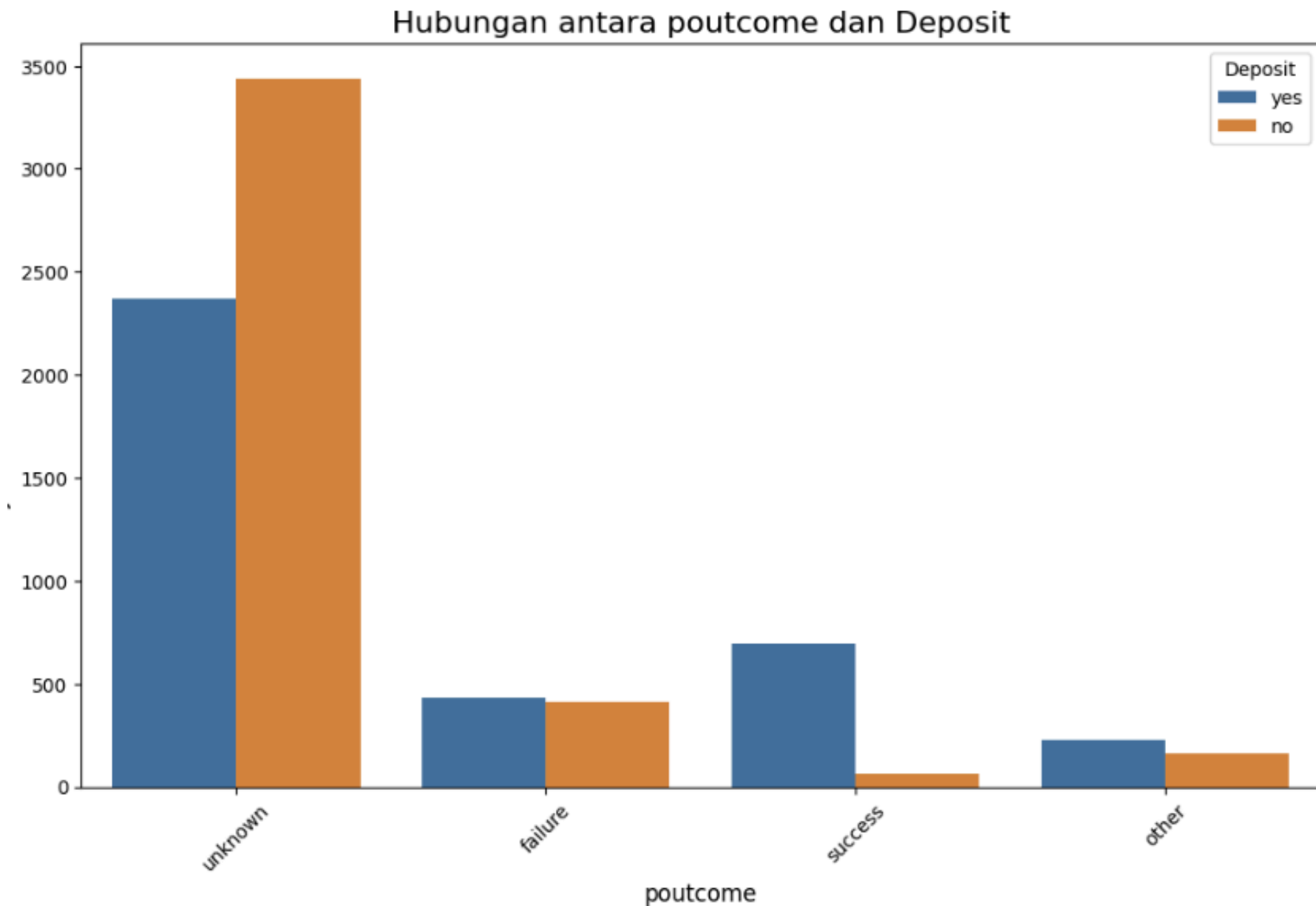
Siapa Nasabah Potensial Kita?

Temuan Kunci dari EDA



Poutcome

Nasabah yang pernah sukses di kampanye sebelumnya hampir pasti akan deposit lagi



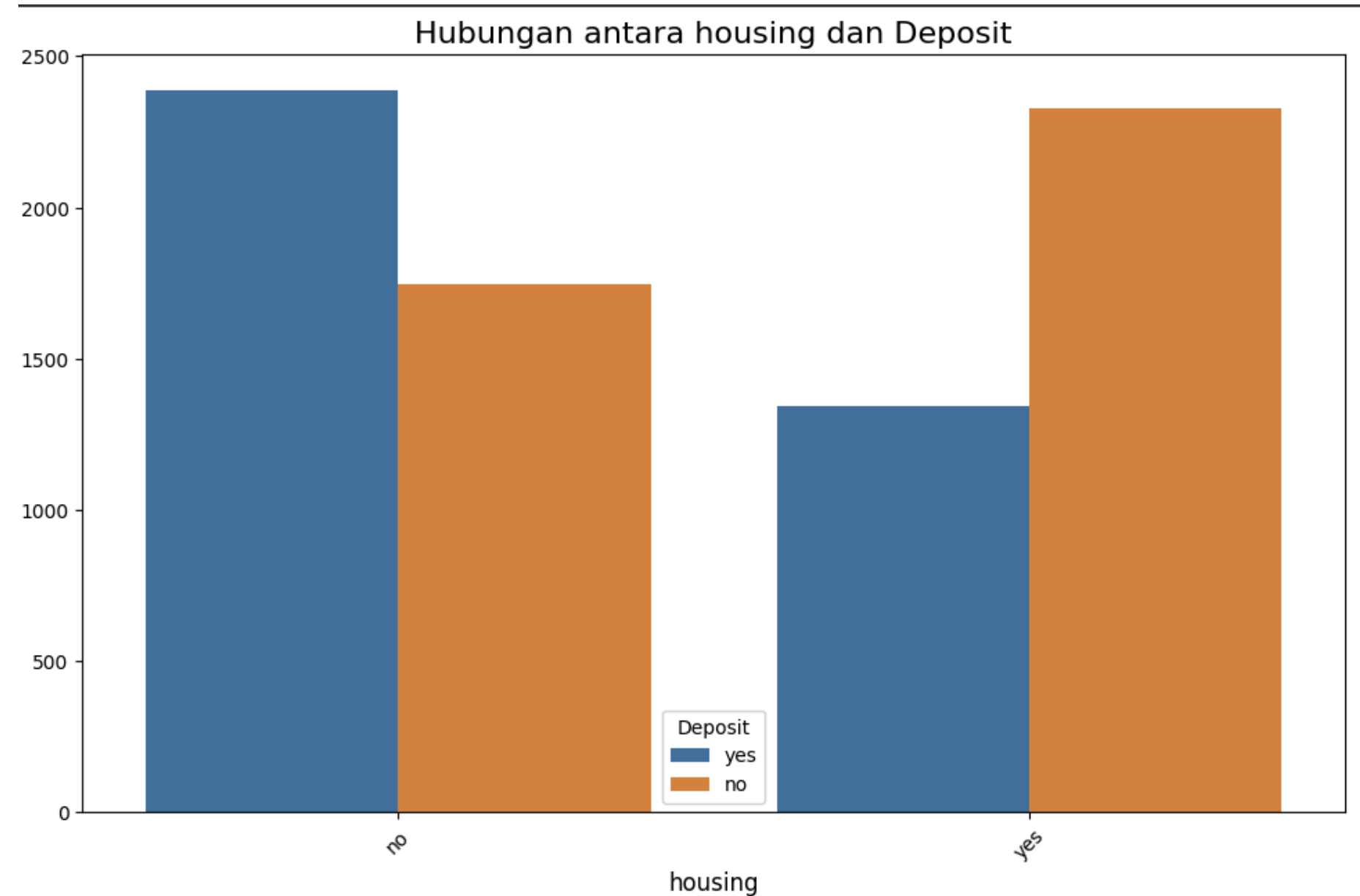
Siapa Nasabah Potensial Kita?

Temuan Kunci dari EDA



Housing

Nasabah yang tidak memiliki kredit rumah atau pinjaman pribadi lebih cenderung untuk deposit.



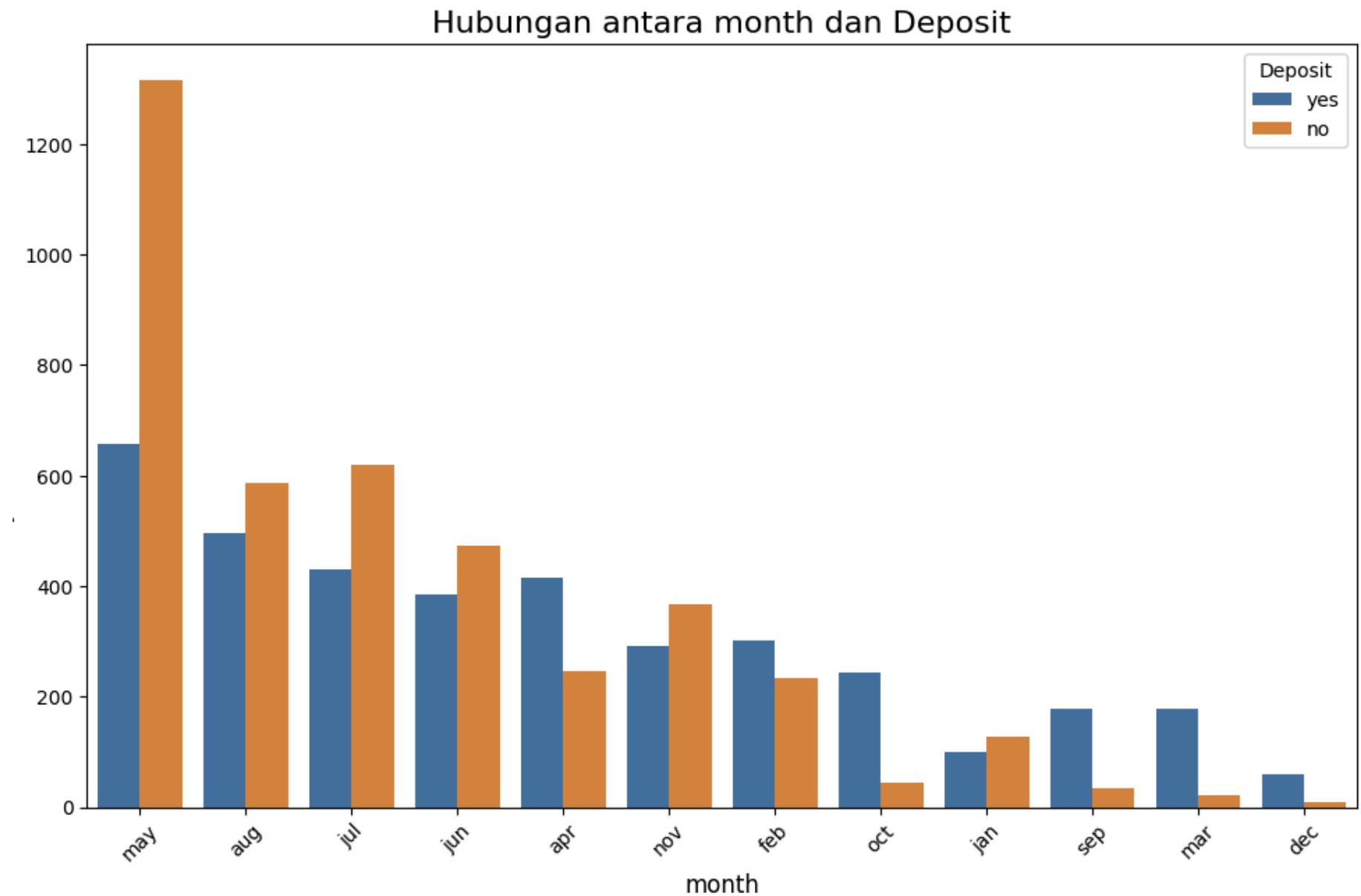
Siapa Nasabah Potensial Kita?

Temuan Kunci dari EDA



Month

Tingkat keberhasilan kampanye sangat tinggi pada bulan Maret, September, Oktober, dan Desember.





Membersihkan dan Menyiapkan Data untuk Model

Persiapan Data

Handling Duplicated Value

Feature Selection

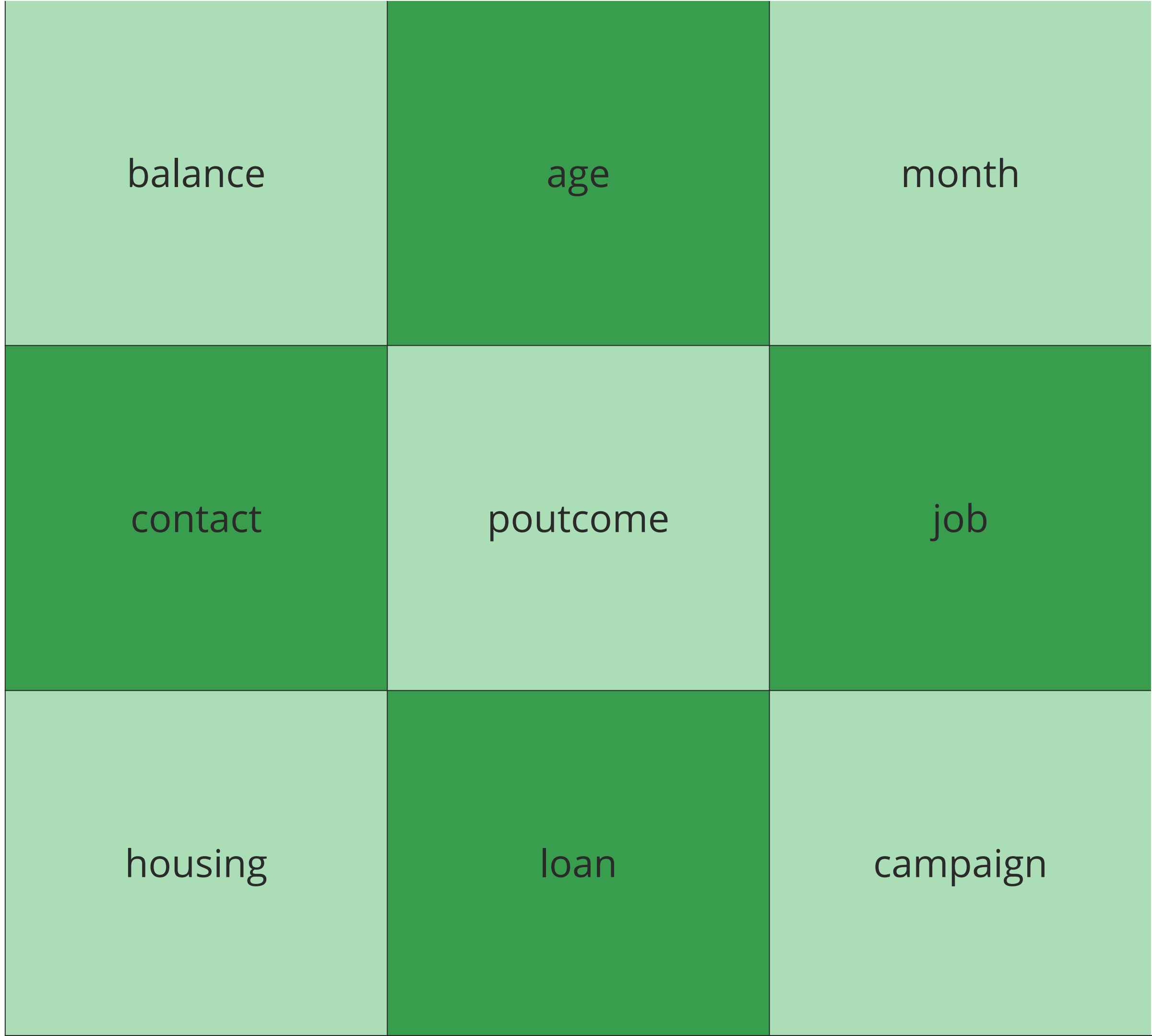
Encoding for categorical features

Normalization for numerical features

Data Splitting

Fokus pada Fitur yang Paling Berpengaruh

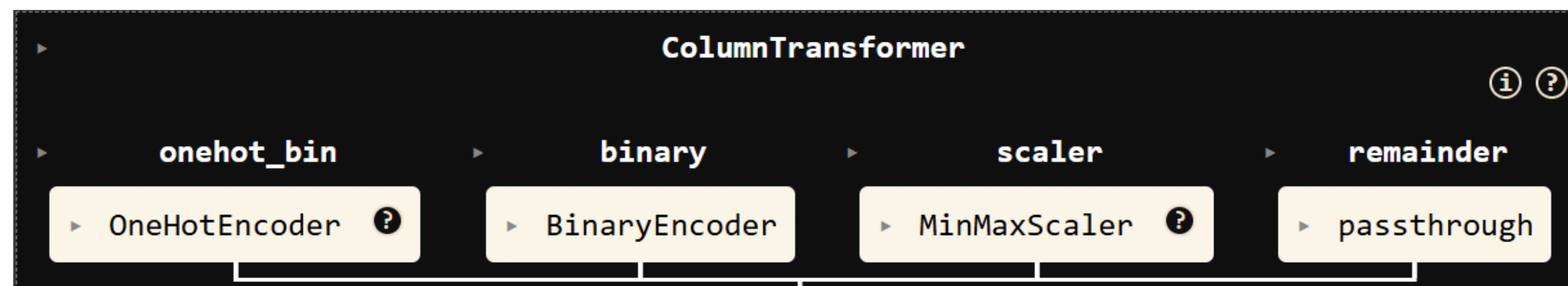
Pemilihan Fitur



Transformasi fitur yang dipilih

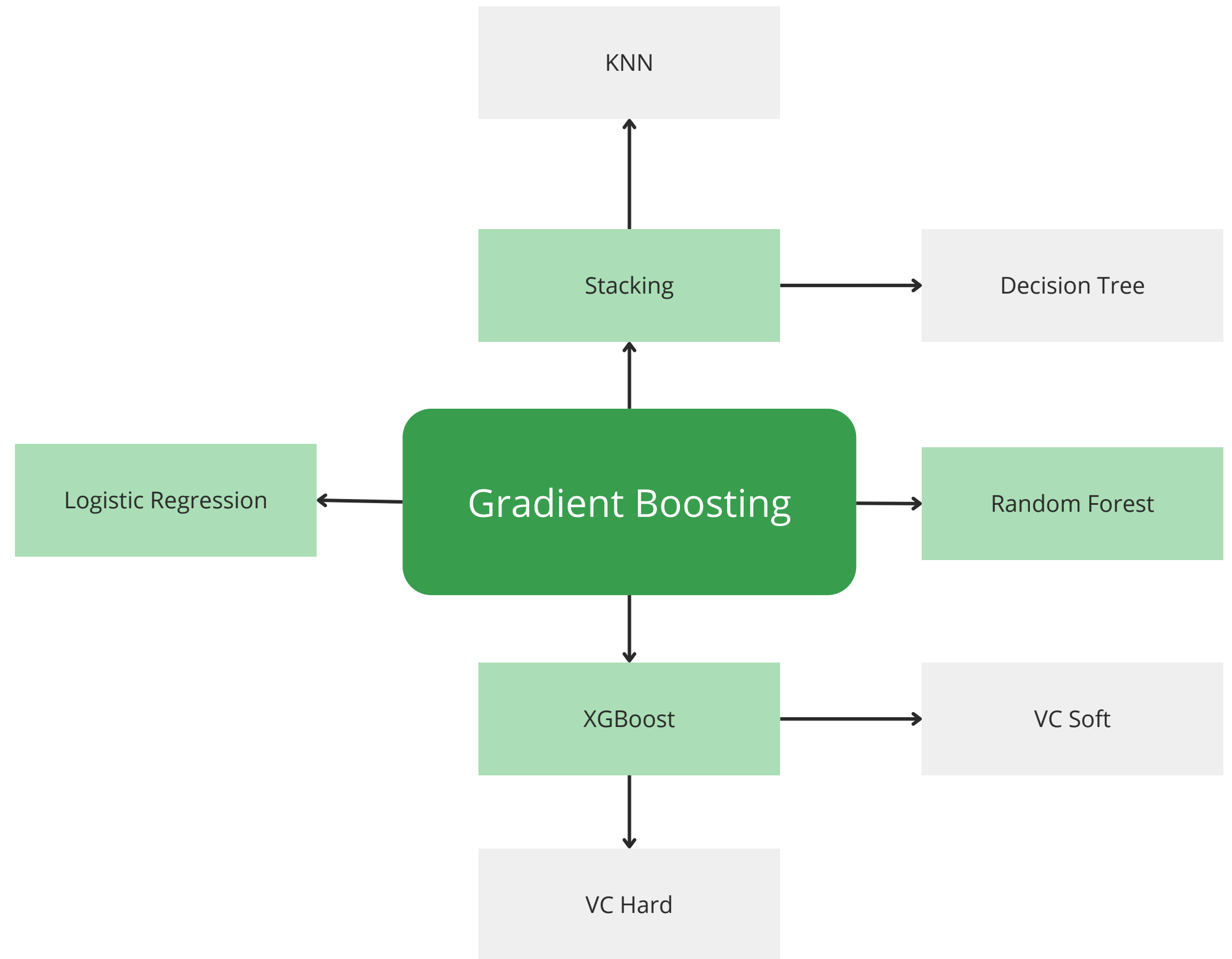
Transformasi Fitur

Sebelum data dapat digunakan oleh model, data mentah yang memiliki format beragam perlu diubah menjadi format numerik yang seragam. Untuk itu, digunakan ColumnTransformer yang bekerja seperti 'manajer proyek' untuk data. Seperti yang diilustrasikan pada gambar, proses ini secara cerdas memisahkan kolom berdasarkan jenisnya: fitur kategorikal dengan sedikit pilihan seperti **month** diubah menggunakan **One-Hot Encoding**; fitur **job** yang memiliki banyak kategori diubah secara efisien dengan **Binary Encoding**; dan fitur numerik seperti **age** dan **balance** disamakan skalanya ke rentang 0-1 menggunakan **Min-Max Scaler**. Hasil akhirnya adalah sebuah dataset yang sepenuhnya numerik, bersih, dan siap untuk tahap pemodelan.



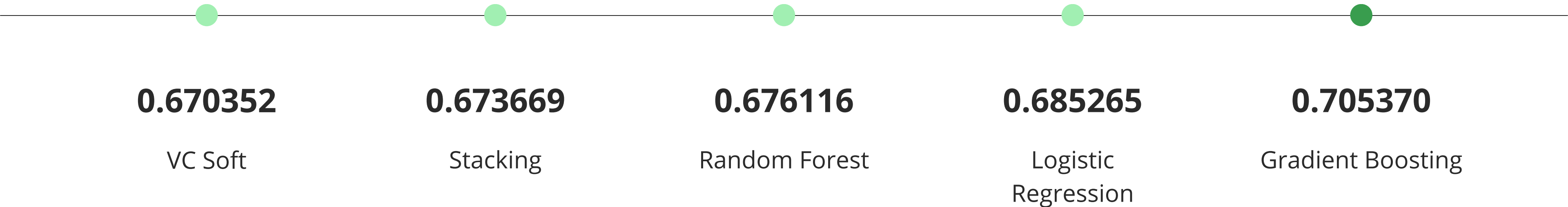
Membangun Model Prediksi

Pendekatan Pemodelan



Model Mana yang Terbaik?

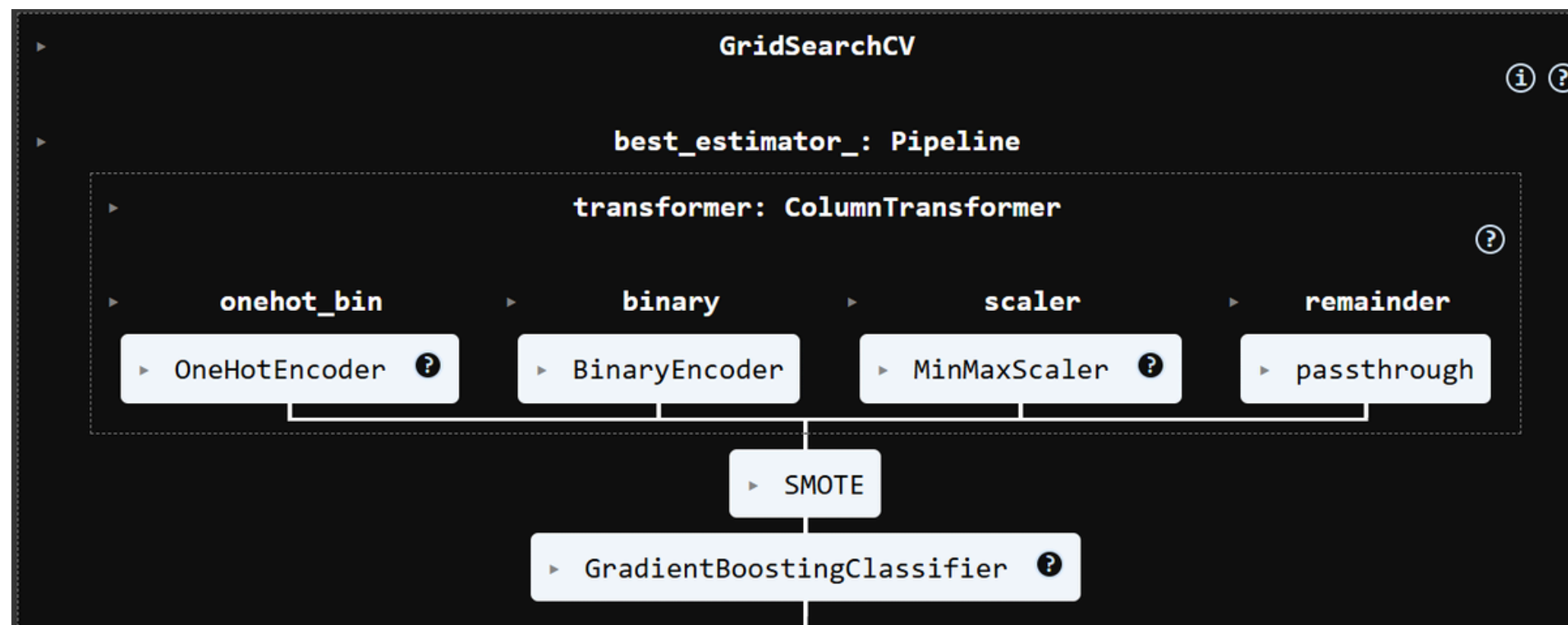
Hasil Evaluasi Model



Kesimpulan Proyek

Kesimpulan **Arsitektur Pipeline Model Final**

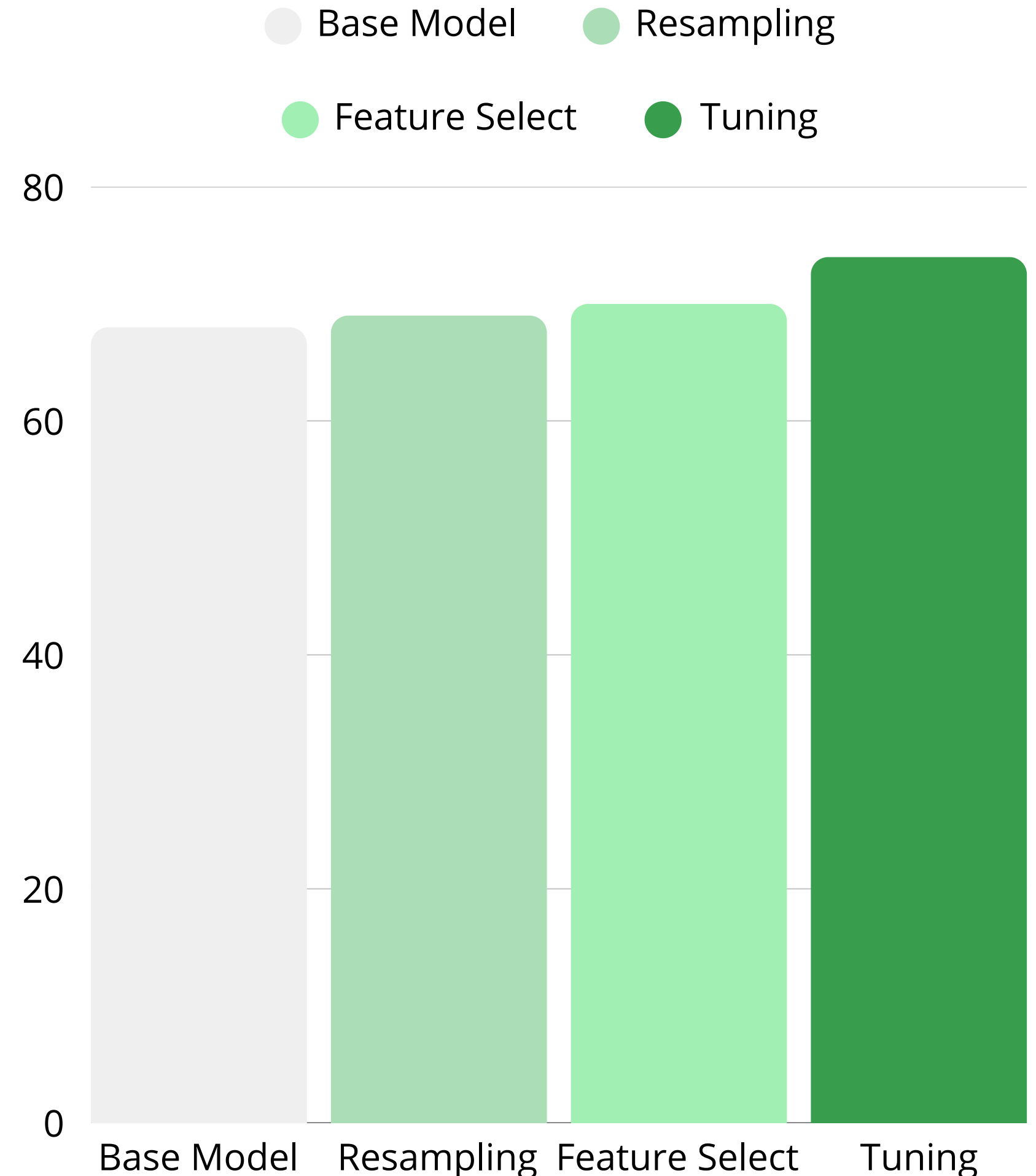
Gambar ini mengilustrasikan arsitektur **pipeline model final**, yang merupakan **sebuah alur kerja dari awal hingga akhir**. Saat data baru masuk, pipeline pertama-tama akan melakukan seluruh tahap preprocessing secara otomatis melalui ColumnTransformer, yang mencakup proses **encoding pada fitur kategorikal** dan **scaling pada fitur numerik**. Setelah itu, data yang telah bersih akan dilewatkan ke tahap **SMOTE** untuk **penyeimbangan kelas**, yang terbukti krusial untuk performa. Terakhir, data yang sudah siap akan dimasukkan ke dalam model **Gradient Boosting** yang telah dioptimalkan untuk menghasilkan prediksi akhir. Dengan adanya pipeline ini, proses prediksi menjadi sangat efisien, konsisten, dan dapat diandalkan untuk data baru di masa depan.



Kesimpulan Proyek

Kesimpulan **Pemilihan Model**

Model **Gradient Boosting** terbukti menjadi yang paling unggul dan terpilih sebagai model final. Peningkatan performa paling signifikan terjadi setelah **penerapan teknik resampling** untuk mengatasi data yang tidak seimbang, yang berhasil menaikkan skor tes terbaik dari **0.686 (baseline)** menjadi **0.694**. Performa kemudian disempurnakan lebih lanjut hingga mencapai skor puncak **0.705** dengan menambahkan **feature selection** dan mencapai skor optimal dan stabil setelah melakukan **Tuning** sebesar **0.740**. Model ini dipilih tidak hanya karena mencapai skor kuantitatif tertinggi, tetapi juga karena konsisten menunjukkan performa yang kuat di berbagai skenario, membuktikan kemampuannya dalam menangani dataset ini secara efektif untuk menjawab permasalahan bisnis.



Kesimpulan Proyek

Kesimpulan **Classification Report**

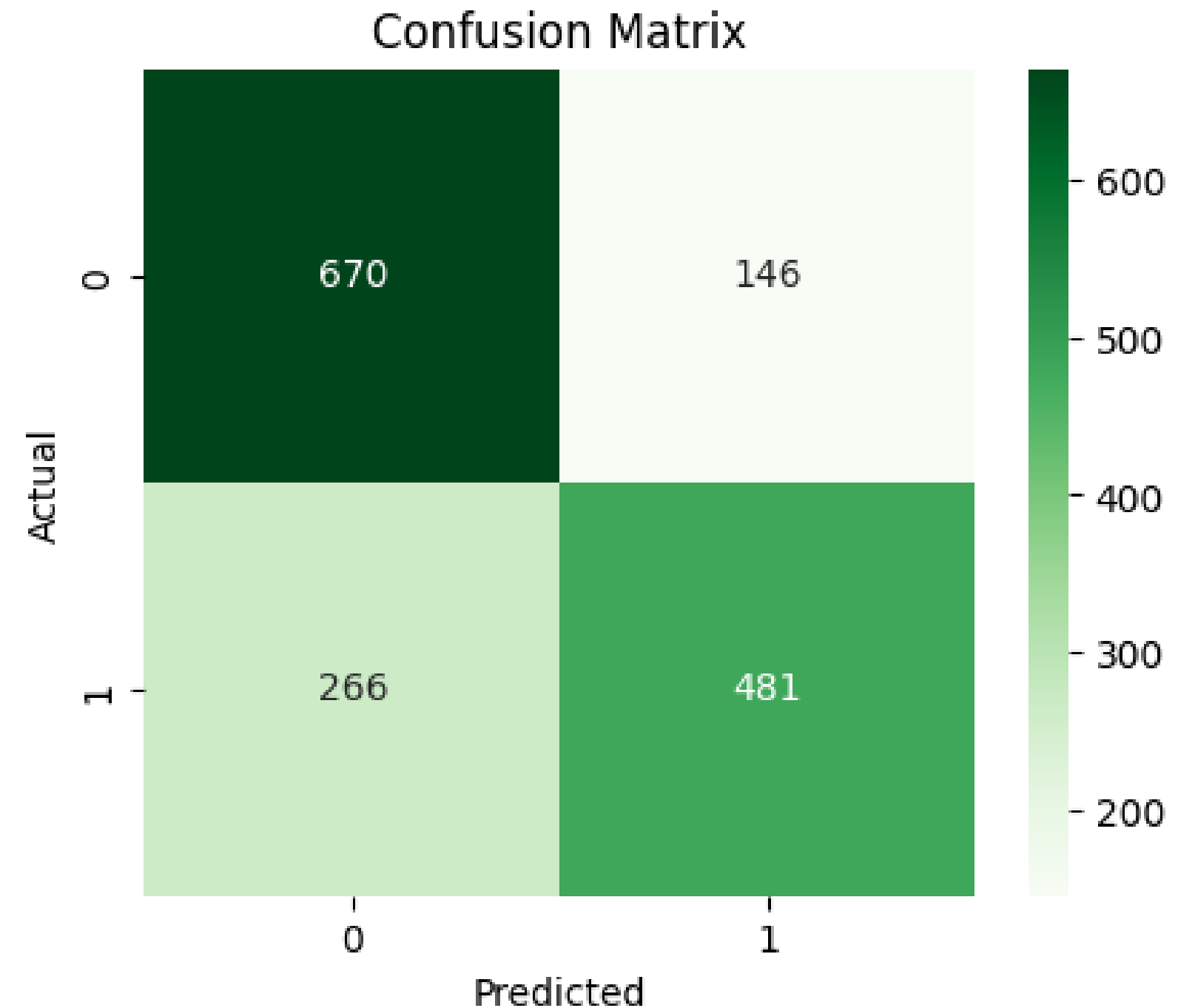
Sebagai kesimpulan, proyek ini telah berhasil membangun model machine learning menggunakan algoritma **Gradient Boosting** yang mampu **memprediksi nasabah potensial** dengan baik, dibuktikan dengan skor rata-rata **Recall 74%**. Metrik **Recall** dipilih sebagai **fokus utama karena dari sisi bisnis**, kehilangan nasabah potensial (*False Negative*) jauh lebih merugikan daripada menghubungi nasabah yang ternyata tidak tertarik (*False Positive*). Analisis juga mengonfirmasi bahwa faktor-faktor seperti **riwayat kampanye nasabah (poutcome)**, **status kepemilikan rumah (housing)**, dan **kondisi finansial (balance)** adalah **prediktor kunci**. Implementasi model ini berpotensi besar untuk mentransformasi kampanye pemasaran dari pendekatan massal menjadi pendekatan berbasis data yang lebih presisi.

	precision	recall	f1-score	support
0	0.72	0.82	0.76	816
1	0.77	0.64	0.70	747
accuracy			0.74	1563
macro avg	0.74	0.73	0.73	1563
weighted avg	0.74	0.74	0.73	1563

Kesimpulan Proyek

Kesimpulan **Confusion Matrix**

Confusion matrix menunjukkan bahwa model sudah mampu mengidentifikasi banyak nasabah yang benar-benar akan melakukan deposit (**True Positive cukup tinggi**) sekaligus mempertahankan akurasi yang baik dalam mengenali nasabah yang tidak akan deposit. Artinya, model sudah memberikan dasar yang kuat untuk **membantu tim pemasaran menargetkan calon nasabah** potensial dengan lebih tepat, dan performanya masih bisa ditingkatkan agar peluang keberhasilan kampanye semakin besar.

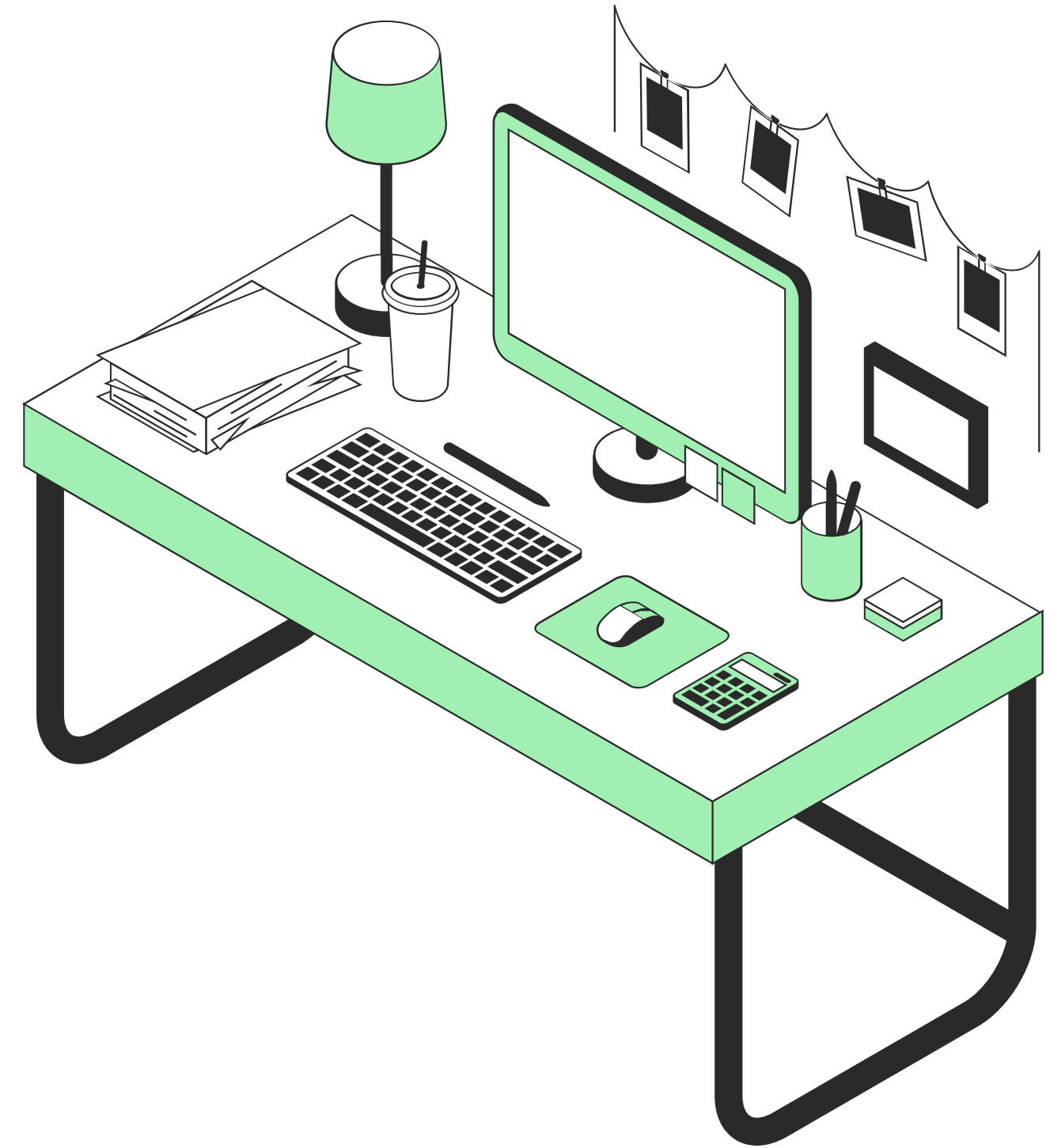


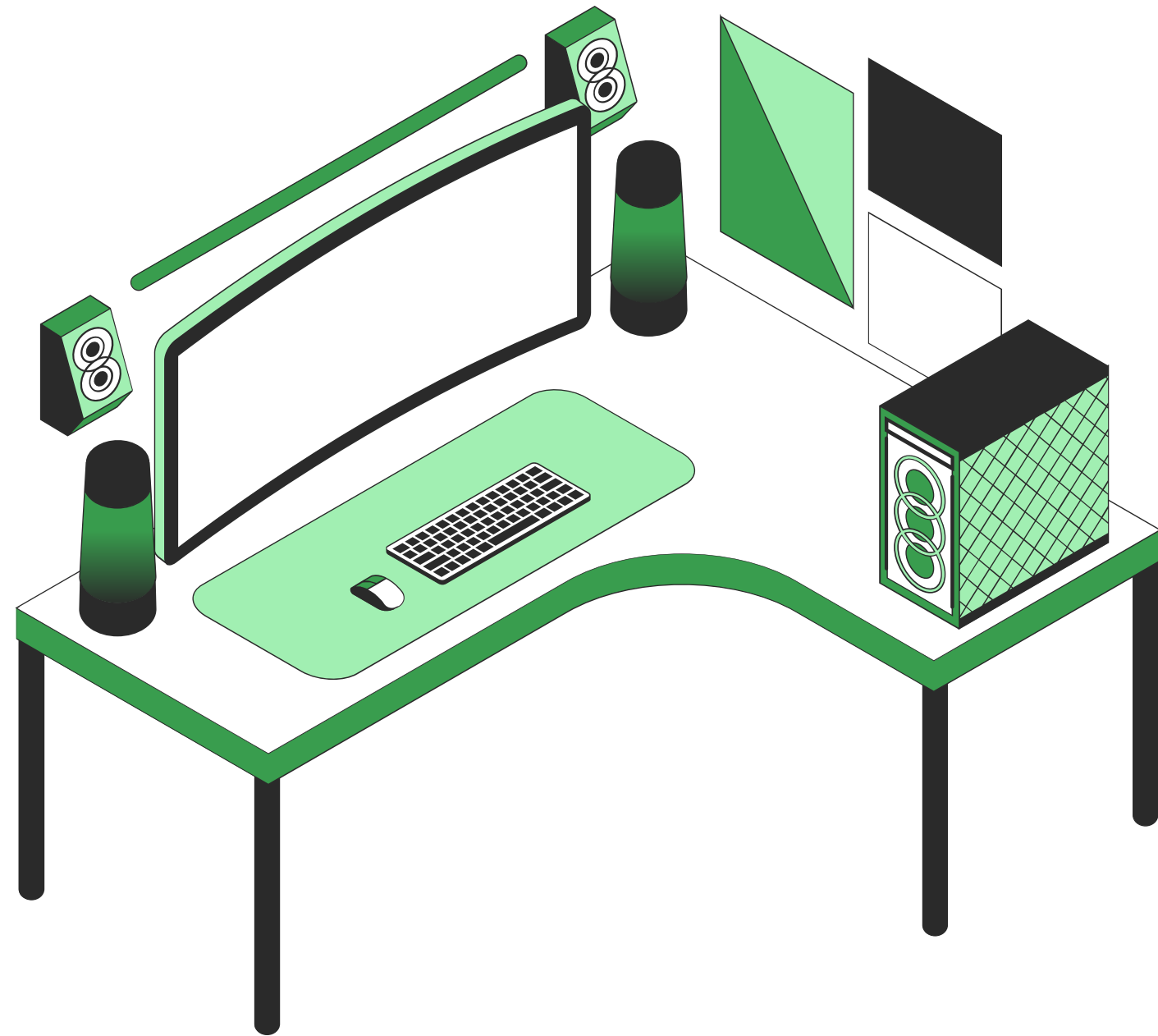
Langkah Selanjutnya?

Rekomendasi Bisnis

Berdasarkan hasil analisis, ada **tiga** rekomendasi tindakan utama.

- **Pertama**, gunakan model ini untuk melakukan scoring dan memprioritaskan nasabah yang akan dihubungi.
- **Kedua**, optimalkan jadwal kampanye dengan mengintensifkan upaya pada bulan-bulan terbukti efektif seperti Maret dan Oktober.
- **Ketiga**, kembangkan penawaran yang lebih menarik untuk segmen nasabah yang tidak memiliki kredit perumahan, karena mereka terbukti lebih reseptif terhadap produk deposito.





Do you
have any
questions?

Terimakasih Banyak!