

NPL Credit Scoring

Certified Junior Data Scientist Program



Project Background

- Bank A adalah sebuah bank yang **ingin melakukan analisa pengajuan kredit** dari nasabahnya, dan Bank A menyediakan data penggunaan kartu kredit nasabah Bank A untuk tujuan tersebut.
- Dengan data tersebut, **Bank A ingin agar kita menciptakan sistem penilaian kelayakan kredit secara otomatis** yang bisa memberikan rekomendasi menerima atau menolak pengajuan kredit nasabah, di mana rekomendasi ini dibuat melalui **prediksi kita tentang potensi kredit macet** dari nasabah tersebut.
- Selain itu, karena ini adalah kasus *credit scoring*, maka jika kita salah **memprediksi seharusnya macet menjadi tidak macet**, maka itu akan **menyebabkan kerugian yang besar pada Bank A** sehingga hal itu hendaknya dihindari.

Problem Statement

- **Faktor-faktor apa saja** yang dapat memprediksi nasabah yang akan kredit macet?
- Apakah ada **behavior berbeda** yang dapat membantu kita untuk mengidentifikasi calon nasabah akan kredit macet?
- Apakah nasabah yang akan kredit macet dapat **diantisipasi** melalui **data historisnya**?



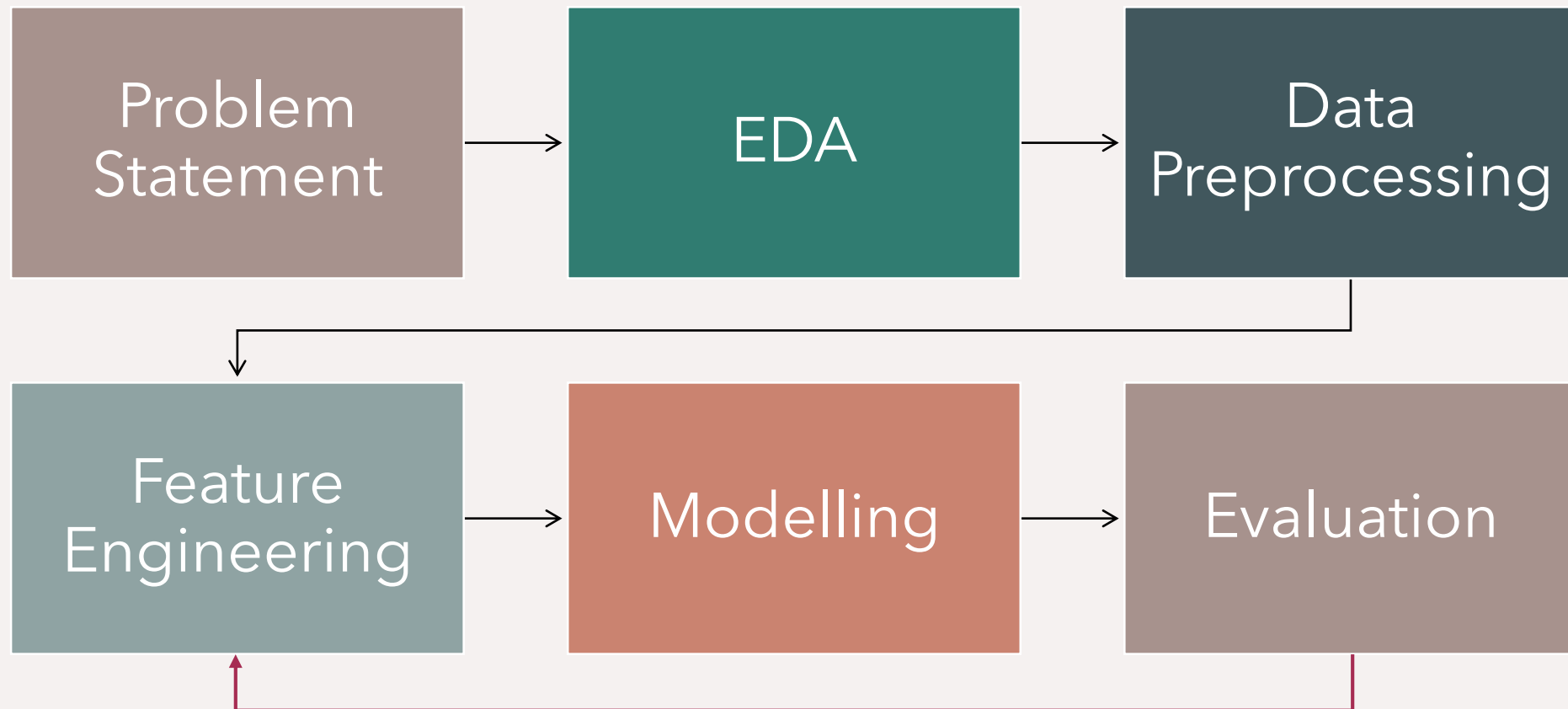
Problem Statement

Secara konkret, masalah kita adalah:

- Kita memiliki problem **classification**
- Kita perlu memprediksi variabel target **flag_kredit_macet**
- Goal kita adalah **memprediksi dengan meminimalisasi false negative** juga, aspek tersebut dapat dinilai melalui metrik **recall**.



Flow Pembuatan Project



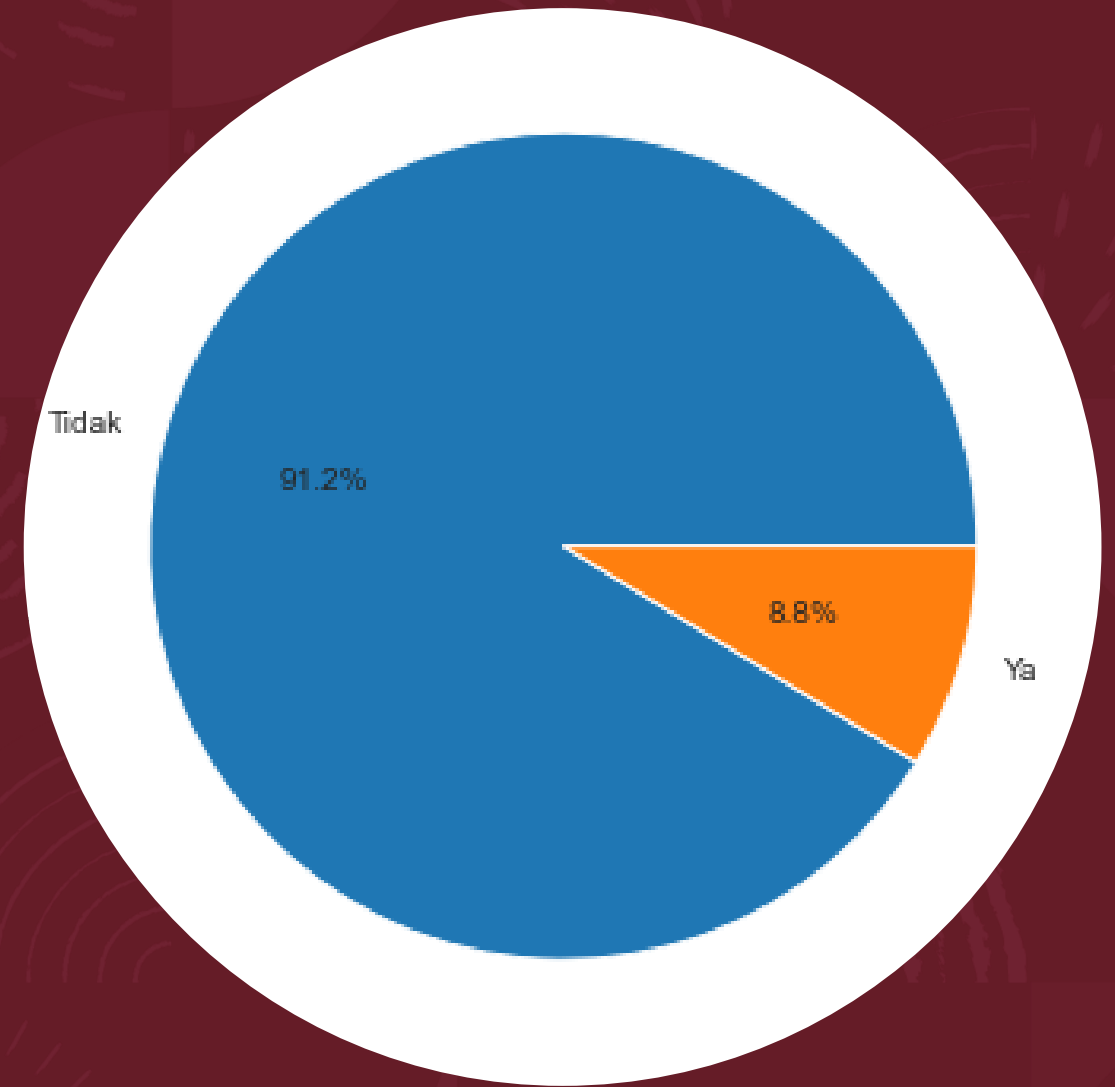


Bagaimana Cara Mendeteksi Nasabah Yang Akan Kredit Macet?



Berapa banyak nasabah kredit macet?

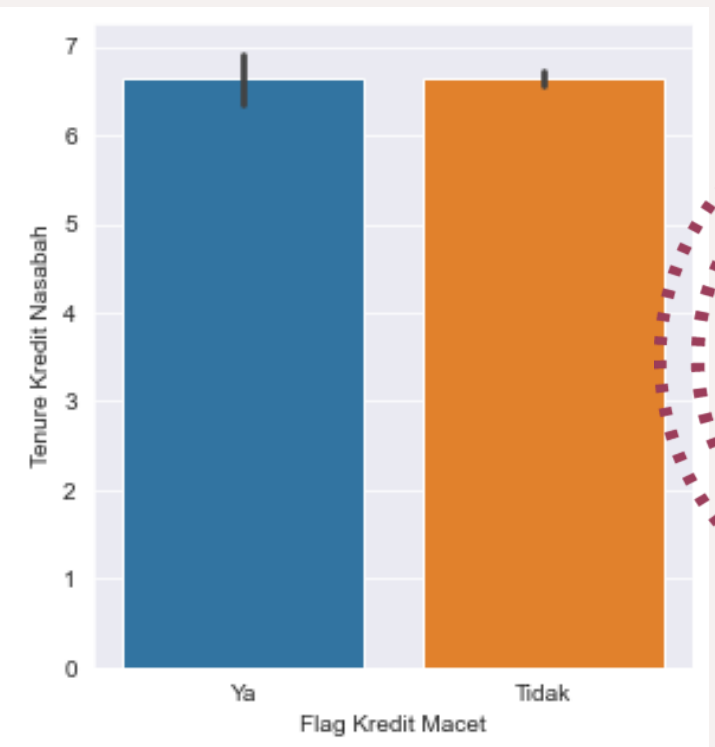
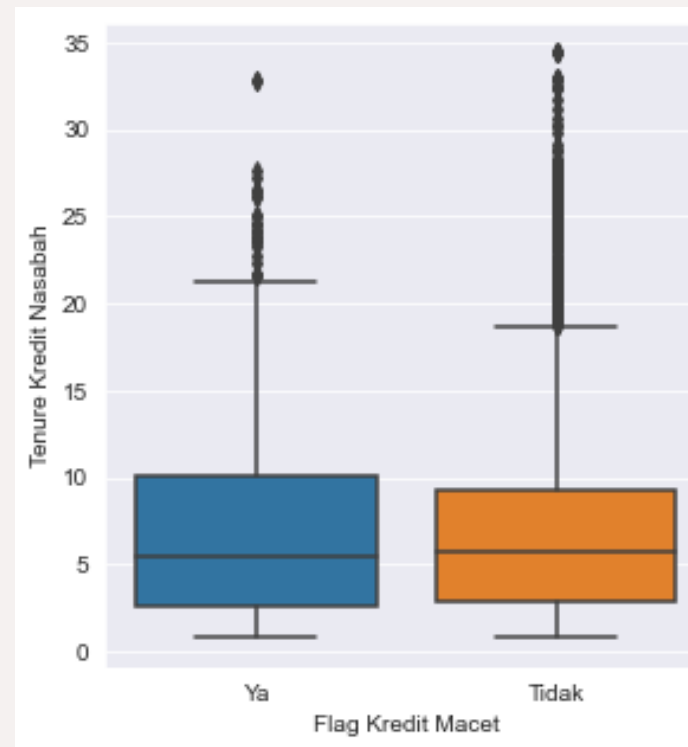
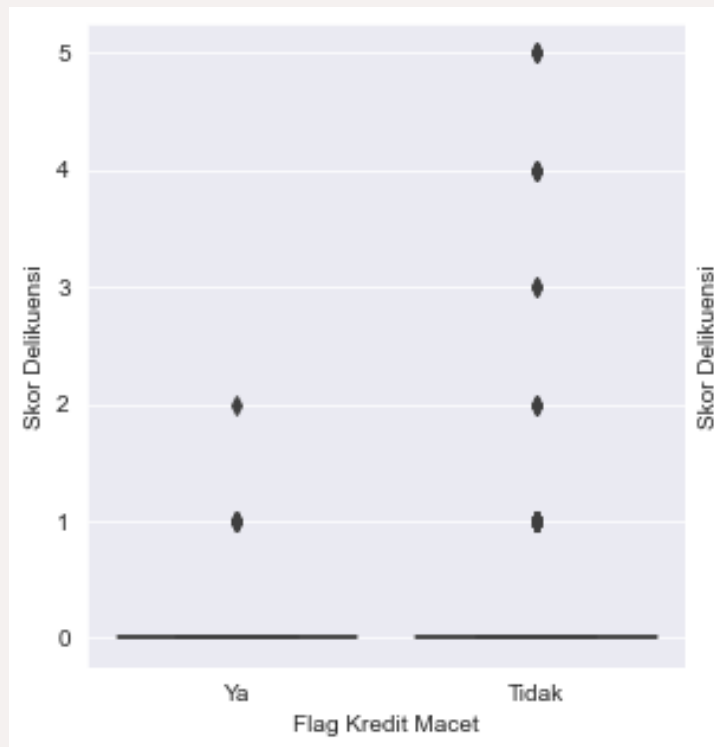
Nasabah kredit macet merupakan **<10%** dari **nasabah** secara keseluruhan, sehingga ini menyebabkan **masalah imbalanced classification** nanti.

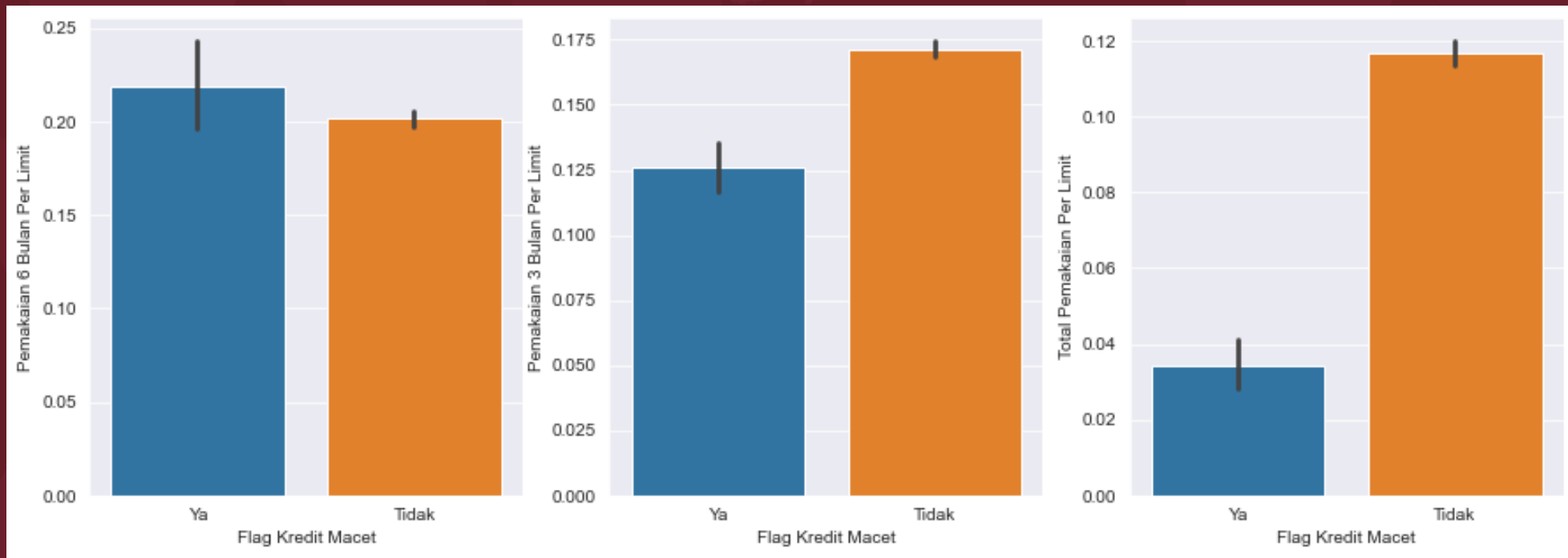


Track Record Historis Biasa...

Tidak mampu memprediksi nasabah kredit macet

- **Sejarah delikueni** sebelumnya tidak dapat memprediksi nasabah kredit macet
- **Sejarah track record** berpuluh-puluh tahun **kredit** bersama Bank A juga tidak dapat memprediksi nasabah kredit macet!

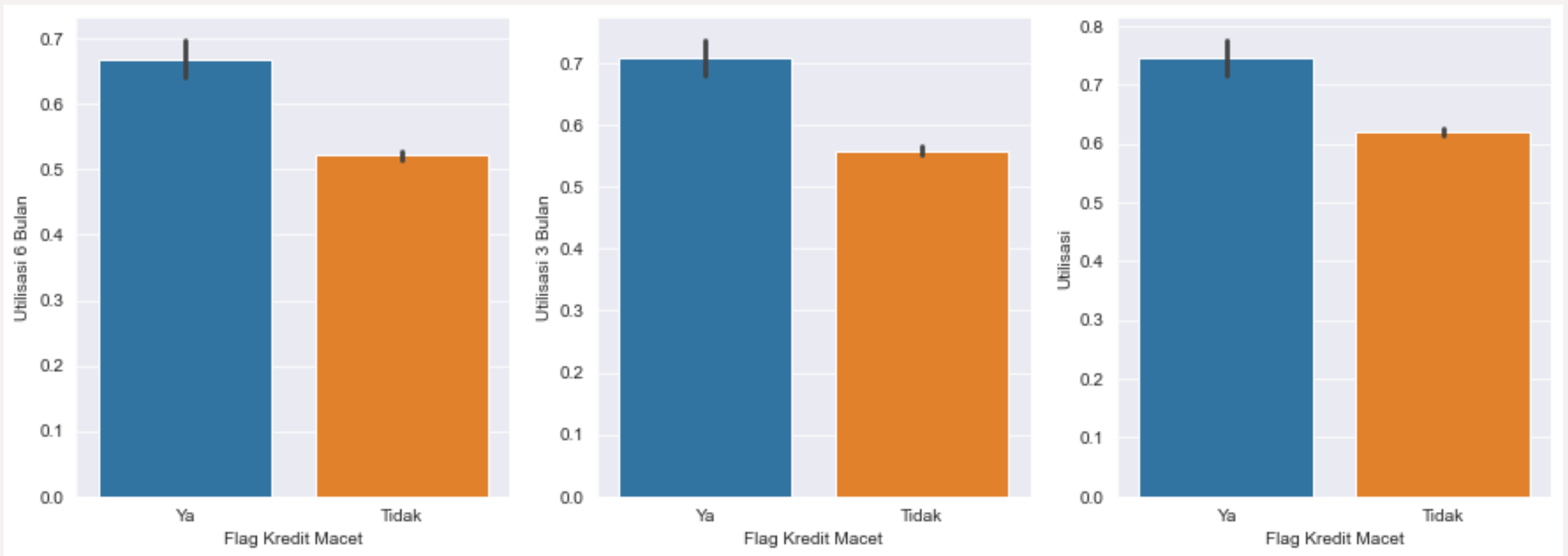




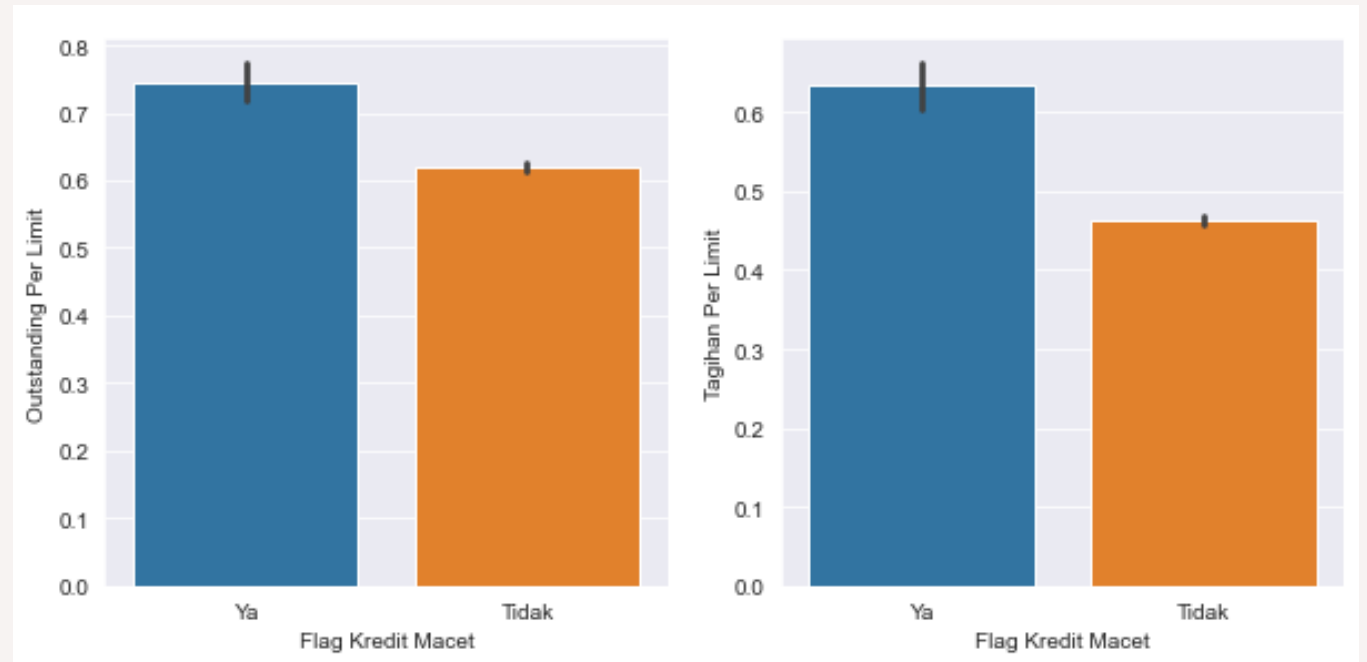
**Dimulai dari pemakaian
6 bulan yang lalu....**

- Periode **6 bulan yang lalu**, nasabah kredit macet memiliki **pemakaian lebih tinggi** daripada nasabah biasa
- Periode **3 bulan yang lalu**, pemakaian nasabah kredit macet **mulai menurun** di bawah nasabah biasa
- **Kini, pemakaian** nasabah kredit macet **jauh di bawah nasabah biasa!**

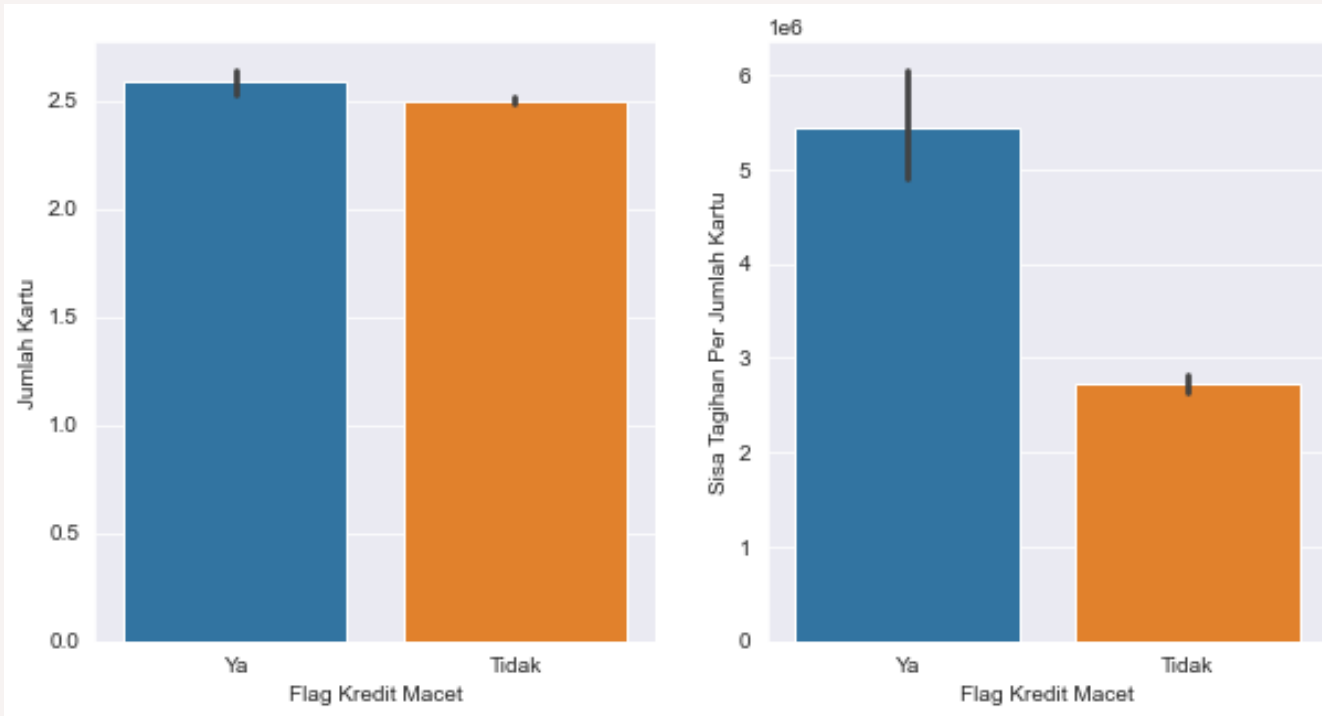
Utilisasi Nasabah Kredit Macet Selalu Tinggi...



Keadaan Kartu Kredit Nasabah Akan Kredit Macet Sekarang (Part 1)



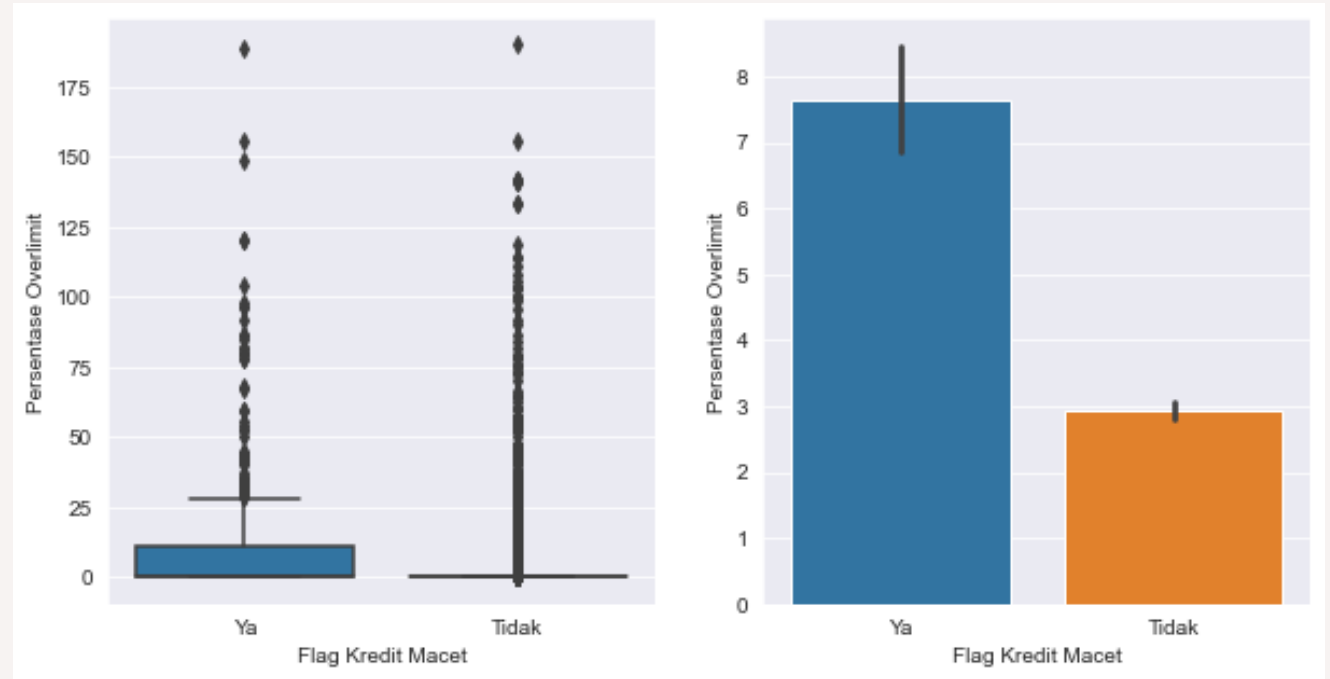
Outstanding dan Tagihan Per Limit yang Tinggi



Keadaan Kartu Kredit Nasabah Akan Kredit Macet Sekarang (Part 2)

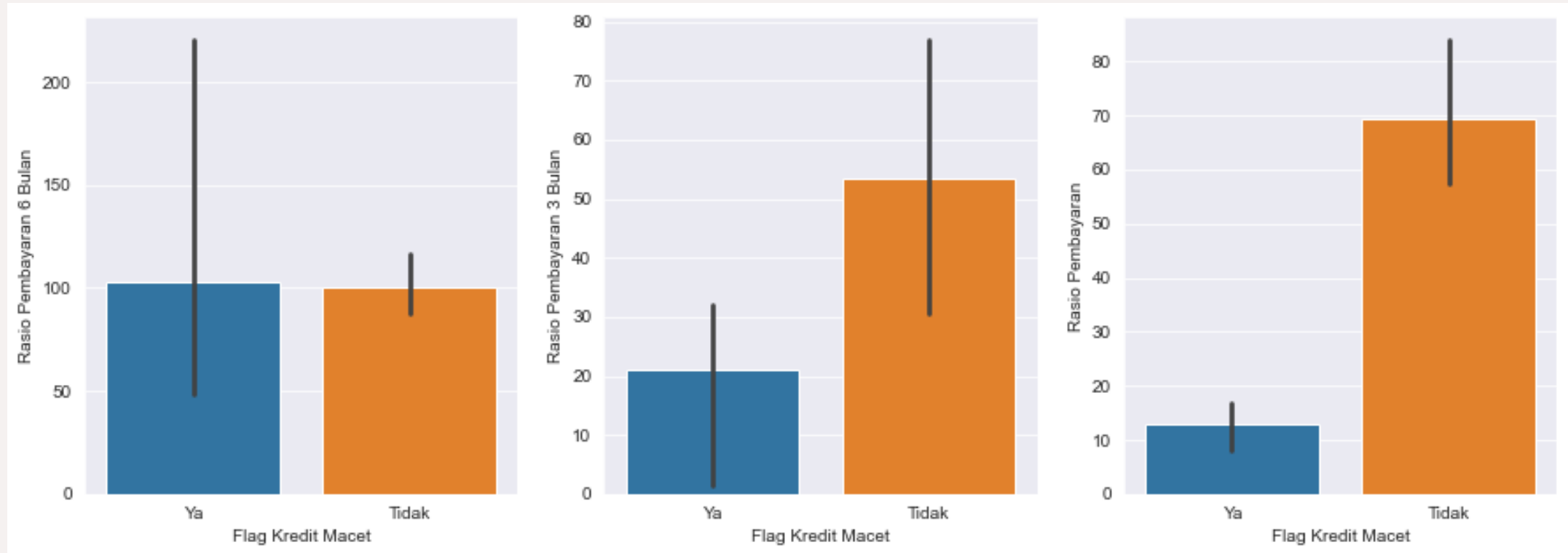
Jumlah Kartu Setara, Sisa Tagihan Jauh Lebih Tinggi

Keadaan Kartu Kredit Nasabah Akan Kredit Macet Sekarang (Part 3)



Kecenderungan Lebih Tinggi untuk Overlimit Kartu

Kunci Nasabah Kredit Macet



Feature Engineering

Feature Selection

Mengeluarkan 3 kolom dengan **korelasi rendah (<0.01)** dan **kurang berpengaruh dari EDA**:

- skor_delikuen
- jumlah_tahun_sejak_pembukaan_kredit
- kode_cabang

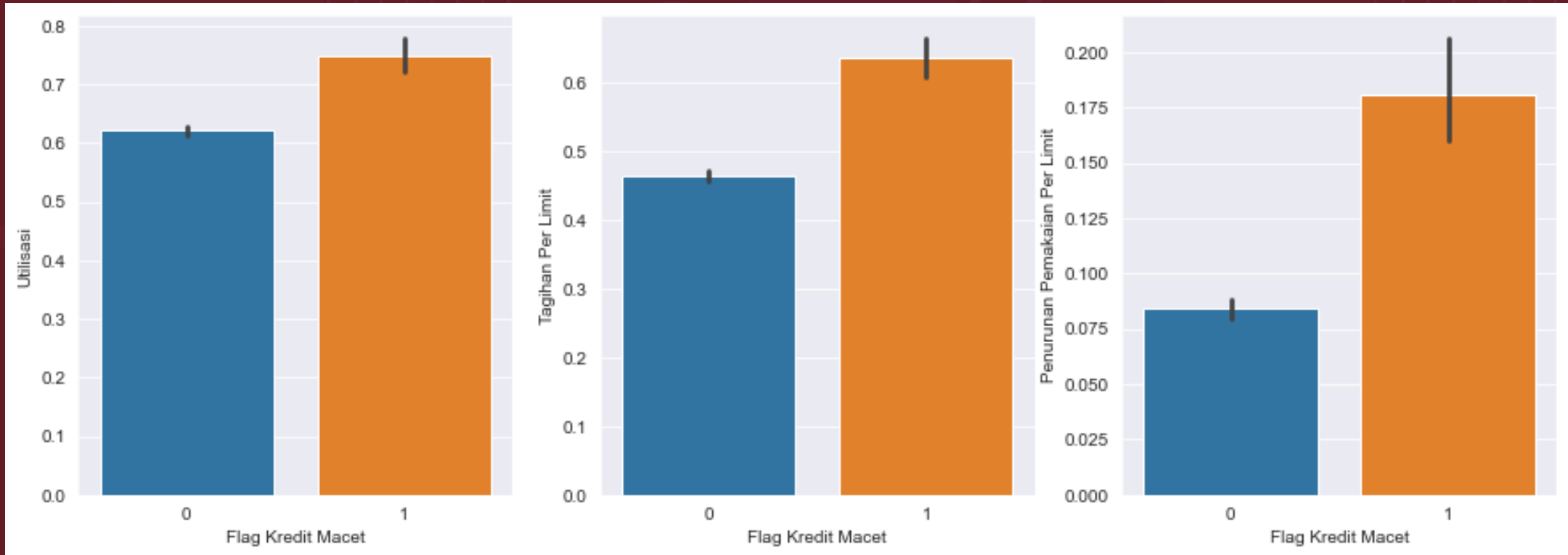
Feature Tambahan

Menambah 3 kolom berdasarkan **hasil eksplorasi saat EDA:**

- utilisasi
- tagihan_per_limit
- penurunan_pemakaian_per_limit

Feature Engineering

Grafik Feature Tambahan



Modelling

- Menggunakan **logistic regression** (mudah diinterpretasi jika diperlukan)
- Data ditangani dengan **Borderline SMOTE** dan (random) **undersampling** untuk menangani imbalanced data
- Model dioptimisasi untuk **recall**, untuk meminimisasi **false negative** yang sangat merugikan Bank A

Hasil Dari Penanganan Imbalanced Class

	Sebelum penanganan	Setelah penanganan
Recall	0.012	0.717
Train Accuracy	0.913	0.775
Validation Accuracy	0.911	0.714



- **C:** 1
- **class_weight:** Balanced
- **penalty:** l2

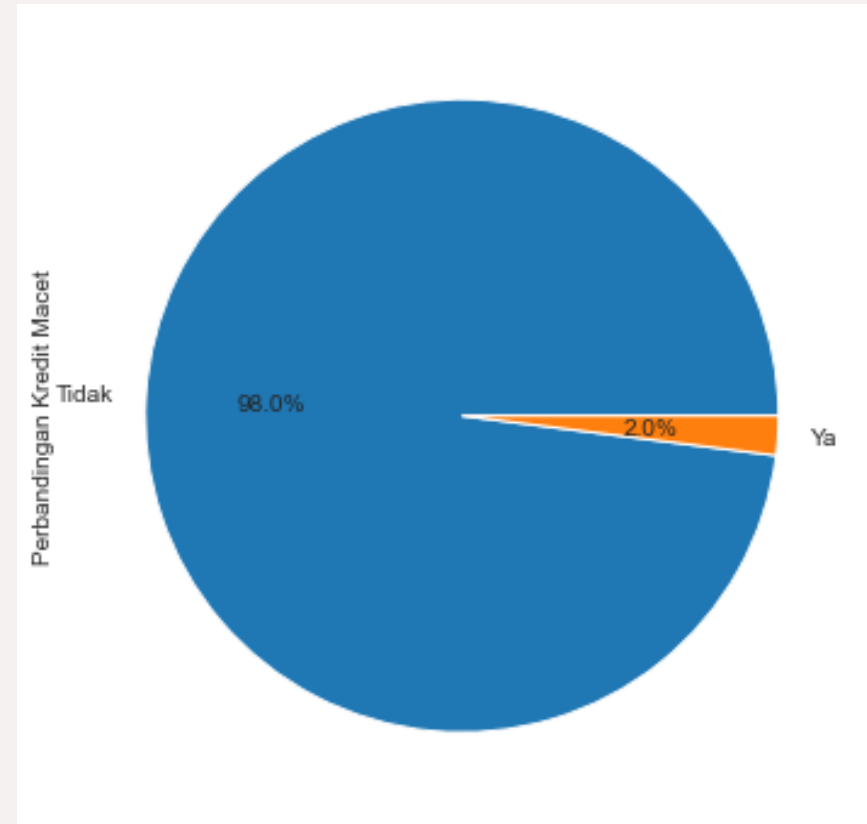
Parameter Akhir

Hasil Akhir

Validation Akhir

Recall	0.796
Train Accuracy	0.762
Test Accuracy	0.656

Prediksi pada data Test



Kesimpulan

- Data memiliki imbalanced classification, di mana hanya ada sekitar 10% nasabah dengan kredit macet sehingga menjadi tantangan dalam pembuatan model.
- Nasabah kredit macet memiliki pattern behaviour yang menarik, di mana umumnya mereka memiliki sisa tagihan tidak terbayar besar sekarang, rasio pembayaran yang terus menurun, serta total pemakaian yang pelan-pelan menurun bersama dengan naiknya tagihan, tetapi tidak ada pola secara skor delikuenensi, jumlah kartu, ataupun *tenure*.
- Masing-masing kolom memiliki korelasi yang relatif kecil dengan variabel target (maksimum hanya 0.2, dan banyak yang jauh lebih kecil daripada itu).
- Secara accuracy, model memiliki akurasi yang sedang-sedang, sekitar 65-75%, namun secara recall yang memang kita optimize, nilai metriknya cukup baik sekitar ~80%.

Saran

- Bank perlu berhati-hati karena faktor-faktor yang biasanya digunakan untuk menilai pengajuan kredit, seperti skor delikuenasi dan *tenure* yang umumnya dilihat sebagai faktor historis, nyaris tidak berdampak dalam kasus ini.
- Nasabah kredit macet dapat diprediksi dan diantisipasi dari behaviournya, sehingga bank dapat melakukan monitoring kepada behaviour nasabah, menggunakan model ini atau teknik lain (misalnya dashboard), untuk mengantisipasi nasabah yang potensial kredit macet dan mengarahkan mereka kembali ke jalan yang benar jika bisa.
- Data dapat diperlengkapi dengan data demografis, data kredit nasabah yang lain selain kartu kredit, aspek 5C lainnya dari nasabah tersebut, di mana setidaknya data demografis pasti dimiliki oleh bank.
- Jika bisa didapatkan data yang lebih lengkap seperti di atas, model bisa dibangun ulang dan ditingkatkan lagi, atau bisa juga crosscheck untuk mengecek misalnya mengapa nasabah kredit macet kerap kali menurun secara rasio pembayaran (apakah *willingly* atau *unwillingly*).



Thank You!