

DESKRIPSI DATASET

❖ Nama dan Sumber Dataset

Nama Dataset : Credit Card Customers

Sumber Dataset : Kaggle

Link : <https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers>

❖ Struktur Data (Skema dan Format)

Dataset ini memiliki **23 fitur (kolom)**, yaitu:

1. **CLIENTNUM** – Nomor unik identifikasi pelanggan (numerik)
2. **Attrition_Flag** – Status pelanggan: *Existing Customer* atau *Attrited Customer* (kategorikal)
3. **Customer_Age** – Usia pelanggan dalam tahun (numerik)
4. **Gender** – Jenis kelamin pelanggan ('M' = laki-laki, 'F' = perempuan) (kategorikal)
5. **Dependent_count** – Jumlah tanggungan pelanggan (numerik)
6. **Education_Level** – Tingkat pendidikan pelanggan (kategorikal)
7. **Marital_Status** – Status pernikahan pelanggan (kategorikal)
8. **Income_Category** – Kategori pendapatan tahunan pelanggan (kategorikal)
9. **Card_Category** – Jenis kartu kredit yang dimiliki pelanggan (kategorikal)
10. **Months_on_book** – Lama hubungan pelanggan dengan bank dalam bulan (numerik)
11. **Total_Relationship_Count** – Jumlah total produk atau layanan bank yang dimiliki pelanggan (numerik)
12. **Months_Inactive_12_mon** – Jumlah bulan tidak aktif dalam 12 bulan terakhir (numerik)
13. **Contacts_Count_12_mon** – Jumlah kontak/interaksi dengan pelanggan dalam 12 bulan terakhir (numerik)
14. **Credit_Limit** – Batas kredit kartu pelanggan (numerik)
15. **Total_Revolving_Bal** – Total saldo bergulir pada kartu kredit (numerik)
16. **Avg_Open_To_Buy** – Rata-rata jumlah kredit yang tersedia untuk digunakan (numerik)
17. **Total_Amt_Chng_Q4_Q1** – Perubahan jumlah transaksi antara Q4 dan Q1 (numerik)
18. **Total_Trans_Amt** – Total jumlah transaksi dalam 12 bulan terakhir (numerik)

19. **Total_Trans_Ct** – Total jumlah transaksi dalam 12 bulan terakhir (numerik)
20. **Total_Ct_Chng_Q4_Q1** – Perubahan jumlah transaksi antara Q4 dan Q1 (numerik)
21. **Avg_Utilization_Ratio** – Rasio rata-rata pemanfaatan kredit pelanggan (numerik)
22. **Naive_Bayes_Classifier_1** – Prediksi churn pelanggan berdasarkan model Naive Bayes (numerik)
23. **Naive_Bayes_Classifier_2** – Prediksi churn kedua dari model Naive Bayes (numerik)

Dataset ini dapat digunakan untuk menganalisis perilaku pelanggan, memprediksi kemungkinan *churn*, dan merancang strategi retensi yang lebih efektif.

Format data yang digunakan adalah CSV, yang umum digunakan untuk penyimpanan data tabular.

Volume Data (Jumlah Rekaman): Dataset ini berisi 10.127 record pelanggan, masing-masing mewakili satu pelanggan kartu kredit.

Dataset ini dapat digunakan untuk berbagai analisis, termasuk prediksi churn pelanggan, segmentasi pelanggan, dan analisis perilaku penggunaan kartu kredit