

Analisis Customer Churn

1. Dataset

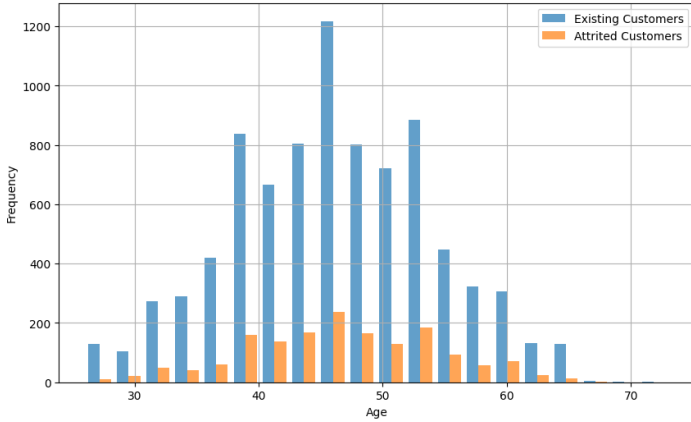
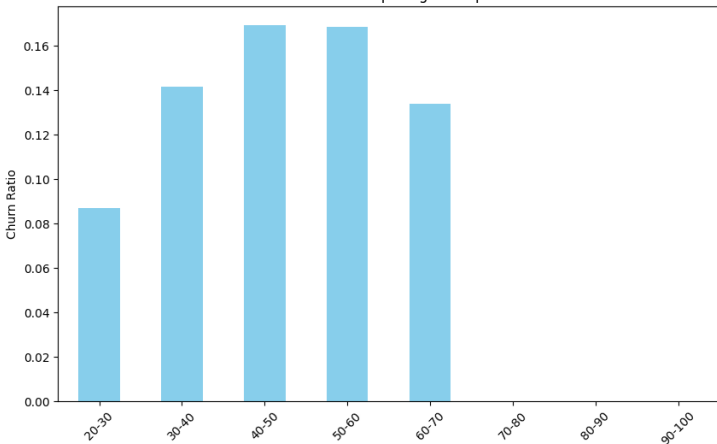
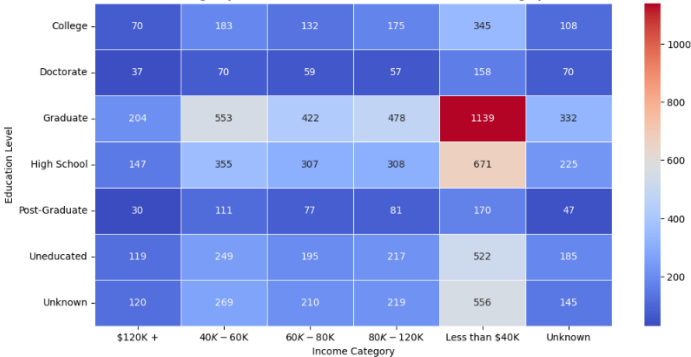
Dataset yang digunakan memiliki fitur-fitur sebagai berikut :

- **user_id**: customer account number.
- **attrition_flag**: customer status (Existing and Attrited).
- **customer_age**: age of the customer.
- **gender**: gender of customer (M for male and F for female).
- **dependent_count**: number of dependents of customers.
- **education_level**: customer education level (Uneducated, High School, Graduate, College, Post-Graduate, Doctorate, and Unknown).
- **marital_status**: customer's marital status (Single, Married, Divorced, and Unknown).
- **income_category**: customer income interval category (Less than \$40K, \$40K-\$60k, \$60K-\$80K, \$80K-\$120K, \$120K +, and Unknown).
- **card_category**: type of card used (Blue, Silver, Gold, and Platinum).
- **months_on_book**: period of being a customer (in months).
- **total_relationship_count**: the number of products used by customers in the bank.
- **months_inactive_12_mon**: period of inactivity for the last 12 months.
- **contacts_count_12_mon**: the number of interactions between the bank and the customer in the last 12 months.
- **credit_limit**: credit card transaction nominal limit in one period.
- **total_revolving_bal**: total funds used in one period.
- **avg_open_to_buy**: the difference between the credit limit set for the cardholder's account and the current balance.
- **total_amt_chng_q4_q1**: increase in customer transaction nominal between quarter 4 and quarter 1.
- **total_trans_amt**: total nominal transaction in the last 12 months.
- **total_trans_ct**: the number of transactions in the last 12 months.
- **total_ct_chng_q4_q1**: the number of customer transactions increased between quarter 4 and quarter 1.
- **avg_utilization_ratio**: percentage of credit card usage.

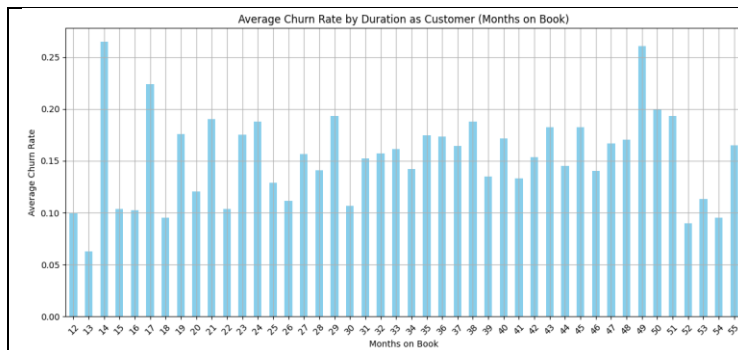
2. Pertanyaan Bisnis

- a. Seperti apa distribusi usia para nasabah? Usia manakah yang cenderung memiliki kemungkinan churn terbesar?
- b. Apakah terdapat korelasi antara tingkat pendidikan dan income category? Bagaimana korelasi dari kedua prediktor tersebut terhadap churn?
- c. Apakah terdapat perbedaan yang signifikan antara gender dengan kemungkinan nasabah untuk churn?
- d. Apakah semakin lama customer menjadi nasabah maka semakin kecil kemungkinan nasabah tersebut untuk churn?
- e. Apakah semakin sering customer berinteraksi dengan bank dalam 1 tahun terakhir mempengaruhi besar atau kecilnya kemungkinan nasabah untuk churn?

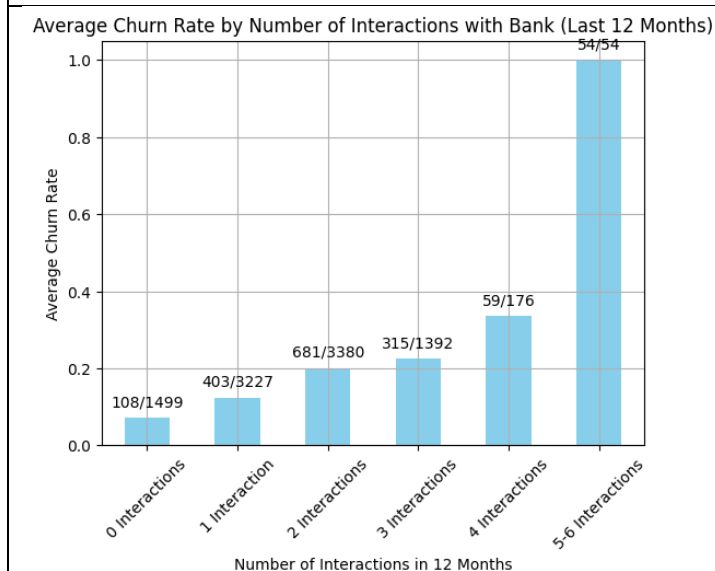
3. Visualisasi

Visualisasi	Penjelasan																																																								
<p>Distribution of Age by Customer Attrition Status</p>  <p>The bar chart displays the frequency of existing customers (blue) and attrited customers (orange) across different age groups. The x-axis represents age from 30 to 70, and the y-axis represents frequency from 0 to 1200. Existing customers show a higher frequency across most age groups, peaking at 45-50. Attrited customers show a higher frequency in the 40-50 age range.</p>	<p>Grafik di samping menunjukkan distribusi usia nasabah berdasarkan status churn(Attrited Customers). Persebaran usia nasabah tersebar cukup merata dengan puncak distribusi pada usia 40-50 tahun. Lalu untuk kasus churn, dapat dilihat bahwa usia menengah (sekitar 40 hingga 50 tahun) cenderung memiliki jumlah churn yang lebih tinggi. Ini mungkin menunjukkan bahwa kelompok usia ini lebih rentan untuk berhenti menjadi nasabah.</p>																																																								
<p>Churn Ratio per Age Group</p>  <p>The bar chart shows the churn ratio for different age groups. The y-axis represents the churn ratio from 0.00 to 0.16. The x-axis represents age groups from 20-30 to 90-100. The churn ratio is highest for the 40-50 and 50-60 age groups, indicating a higher churn rate for middle-aged customers.</p>	<p>Bar Plot di samping menunjukkan rasio churn per kelompok usia, dimana kelompok usia 40-50 tahun dan 50-60 tahun memiliki rasio churn yang tertinggi di antara kelompok usia lainnya. Sedangkan kelompok usia 60-70 tahun dan 70-80 tahun memiliki rasio churn yang lebih rendah, yang berarti nasabah yang lebih tua cenderung lebih sedikit untuk churn dibandingkan dengan yang lebih muda.</p>																																																								
<p>Contingency Table between Education Level and Income Category</p>  <table><tr><th>Education Level</th><th>\$120K +</th><th>40K - 60K</th><th>60K - 80K</th><th>80K - 120K</th><th>Less than \$40K</th><th>Unknown</th></tr><tr><td>College</td><td>70</td><td>183</td><td>132</td><td>175</td><td>345</td><td>108</td></tr><tr><td>Doctorate</td><td>37</td><td>70</td><td>59</td><td>57</td><td>158</td><td>70</td></tr><tr><td>Graduate</td><td>204</td><td>553</td><td>422</td><td>478</td><td>1139</td><td>332</td></tr><tr><td>High School</td><td>147</td><td>355</td><td>307</td><td>308</td><td>671</td><td>225</td></tr><tr><td>Post-Graduate</td><td>30</td><td>111</td><td>77</td><td>81</td><td>170</td><td>47</td></tr><tr><td>Uneducated</td><td>119</td><td>249</td><td>195</td><td>217</td><td>522</td><td>185</td></tr><tr><td>Unknown</td><td>120</td><td>269</td><td>210</td><td>219</td><td>556</td><td>145</td></tr></table>	Education Level	\$120K +	40K - 60K	60K - 80K	80K - 120K	Less than \$40K	Unknown	College	70	183	132	175	345	108	Doctorate	37	70	59	57	158	70	Graduate	204	553	422	478	1139	332	High School	147	355	307	308	671	225	Post-Graduate	30	111	77	81	170	47	Uneducated	119	249	195	217	522	185	Unknown	120	269	210	219	556	145	<p>Berdasarkan tabel korelasi disamping, Kategori "Graduate" memiliki jumlah nasabah terbanyak di berbagai kelompok pendapatan, khususnya pada pendapatan "Less than \$40K". Lalu tingkat pendidikan yang lebih tinggi (seperti "Doctorate" dan "Post-Graduate") menunjukkan distribusi nasabah yang cenderung lebih kecil dibandingkan kategori pendidikan lainnya. Adapun pendapatan kategori "Less than \$40K" memiliki nasabah dari berbagai tingkat pendidikan, menunjukkan inklusivitas pada rentang pendidikan yang luas.</p>
Education Level	\$120K +	40K - 60K	60K - 80K	80K - 120K	Less than \$40K	Unknown																																																			
College	70	183	132	175	345	108																																																			
Doctorate	37	70	59	57	158	70																																																			
Graduate	204	553	422	478	1139	332																																																			
High School	147	355	307	308	671	225																																																			
Post-Graduate	30	111	77	81	170	47																																																			
Uneducated	119	249	195	217	522	185																																																			
Unknown	120	269	210	219	556	145																																																			

<table border="1"><caption>Churn vs Education Level</caption><thead><tr><th>Education Level</th><th>attrition_flag = 0 (Count)</th><th>attrition_flag = 1 (Count)</th></tr></thead><tbody><tr><td>High School</td><td>1700</td><td>300</td></tr><tr><td>Graduate</td><td>2600</td><td>500</td></tr><tr><td>Uneducated</td><td>1250</td><td>250</td></tr><tr><td>Unknown</td><td>1250</td><td>250</td></tr><tr><td>College</td><td>850</td><td>150</td></tr><tr><td>Post-Graduate</td><td>400</td><td>100</td></tr><tr><td>Doctorate</td><td>350</td><td>100</td></tr></tbody></table>	Education Level	attrition_flag = 0 (Count)	attrition_flag = 1 (Count)	High School	1700	300	Graduate	2600	500	Uneducated	1250	250	Unknown	1250	250	College	850	150	Post-Graduate	400	100	Doctorate	350	100	<p>Dari grafik disamping dapat dilihat bahwa nasabah dengan tingkat pendidikan "Graduate" memiliki jumlah churn tertinggi secara absolut, tetapi hal ini juga dikarenakan jumlah total nasabah dalam kelompok ini adalah yang paling besar. Tingkat pendidikan seperti "Doctorate" dan "Post-Graduate" menunjukkan proporsi churn yang lebih tinggi dibandingkan dengan tingkat pendidikan lainnya. Jadi nasabah dengan tingkat pendidikan yang lebih tinggi mungkin memiliki harapan yang lebih tinggi terhadap layanan bank. Jika layanan tersebut tidak memenuhi ekspektasi, kemungkinan churn akan meningkat.</p>
Education Level	attrition_flag = 0 (Count)	attrition_flag = 1 (Count)																							
High School	1700	300																							
Graduate	2600	500																							
Uneducated	1250	250																							
Unknown	1250	250																							
College	850	150																							
Post-Graduate	400	100																							
Doctorate	350	100																							
<table border="1"><caption>Churn vs Income Category</caption><thead><tr><th>Income Category</th><th>attrition_flag = 0 (Count)</th><th>attrition_flag = 1 (Count)</th></tr></thead><tbody><tr><td>60K-80K</td><td>1200</td><td>200</td></tr><tr><td>Less than 40K</td><td>2900</td><td>600</td></tr><tr><td>80K-120K</td><td>1300</td><td>250</td></tr><tr><td>40K-60K</td><td>1500</td><td>300</td></tr><tr><td>\$120K+</td><td>600</td><td>150</td></tr><tr><td>Unknown</td><td>900</td><td>200</td></tr></tbody></table>	Income Category	attrition_flag = 0 (Count)	attrition_flag = 1 (Count)	60K-80K	1200	200	Less than 40K	2900	600	80K-120K	1300	250	40K-60K	1500	300	\$120K+	600	150	Unknown	900	200	<p>Dari grafik disamping dapat dilihat bahwa nasabah dengan kategori pendapatan "Less than \$40K" menunjukkan jumlah churn tertinggi secara absolut. Churn juga terlihat cukup signifikan pada kategori pendapatan yang lebih rendah (e.g., "40K-60K"). Hal ini mungkin karena pendapatan yang rendah membuat nasabah lebih sensitif terhadap biaya layanan atau penawaran produk yang dianggap kurang memberikan nilai tambah.</p>			
Income Category	attrition_flag = 0 (Count)	attrition_flag = 1 (Count)																							
60K-80K	1200	200																							
Less than 40K	2900	600																							
80K-120K	1300	250																							
40K-60K	1500	300																							
\$120K+	600	150																							
Unknown	900	200																							
<table border="1"><caption>Churn vs Gender</caption><thead><tr><th>Gender</th><th>attrition_flag = 0 (Count)</th><th>attrition_flag = 1 (Count)</th></tr></thead><tbody><tr><td>M</td><td>4050</td><td>700</td></tr><tr><td>F</td><td>4400</td><td>950</td></tr></tbody></table>	Gender	attrition_flag = 0 (Count)	attrition_flag = 1 (Count)	M	4050	700	F	4400	950	<p>Dari grafik disamping, kita dapat melihat bahwa nasabah perempuan tampaknya memiliki jumlah churn yang lebih tinggi daripada laki-laki. Ini dapat mengindikasikan bahwa nasabah perempuan lebih cenderung meninggalkan bank dibandingkan dengan nasabah laki-laki. Jadi Gender berpengaruh terhadap kemungkinan churn, karena perbedaan dalam jumlah nasabah yang churn antara perempuan dan laki-laki terlihat jelas.</p>															
Gender	attrition_flag = 0 (Count)	attrition_flag = 1 (Count)																							
M	4050	700																							
F	4400	950																							



Grafik disamping menunjukkan adanya fluktuasi churn rate antar bulan. Meskipun ada kecenderungan penurunan churn rate pada beberapa bulan, namun ada lonjakan churn yang lebih tinggi pada bulan tertentu (misalnya di bulan 13 ke bulan 14 atau dibulan 48 ke 49). Hal ini menunjukkan bahwa tingkat churn tidak selalu mengikuti pola yang konsisten. Dengan kata lain, tidak selalu benar bahwa semakin lama nasabah menjadi pelanggan, semakin kecil kemungkinan mereka untuk churn.



Dari grafik disamping dapat diketahui bahwa frekuensi interaksi dengan bank tidak selalu berkorelasi dengan penurunan churn rate. Sebaliknya, meskipun lebih sering berinteraksi, banyak kelompok nasabah yang mengalami peningkatan churn rate, terutama pada kelompok yang memiliki lebih dari 2 interaksi.

4. Performa Model

Model yang digunakan : RandomForestClassifier

Hasil Evaluasi :



Classification Report

Classification report memberikan informasi tentang kinerja model untuk kedua kelas (0: Existing Customer, 1: Attrited Customer).

Penjelasan Tiap Metrik

1. Precision: Precision menunjukkan seberapa sering model benar dalam memprediksi suatu kelas.
 - Untuk Existing Customer (0): Dari semua pelanggan yang diprediksi sebagai Existing Customer, 95% benar.
 - Untuk Attrited Customer (1): Dari semua pelanggan yang diprediksi sebagai Attrited Customer, 94% benar.
2. Recall: Recall menunjukkan seberapa banyak pelanggan yang sebenarnya termasuk dalam suatu kelas yang berhasil diidentifikasi oleh model.
 - Untuk Existing Customer (0): Dari semua pelanggan yang sebenarnya Existing Customer, 99% diidentifikasi dengan benar.
 - Untuk Attrited Customer (1): Dari semua pelanggan yang sebenarnya churn, hanya 76% diidentifikasi dengan benar. (Catatan: Recall untuk churn relatif rendah.)
3. F1-Score: Kombinasi antara precision dan recall, memberikan gambaran keseimbangan keduanya.
 - Untuk Existing Customer (0): 0.97 menunjukkan kinerja yang hampir sempurna.
 - Untuk Attrited Customer (1): 0.84 menunjukkan kinerja yang baik, tetapi ada ruang untuk perbaikan karena recall yang lebih rendah.
4. Support: Jumlah sampel di setiap kelas.
 - Existing Customer (0): 1699.
 - Attrited Customer (1): 327. (Kelas ini lebih kecil, yang dapat menyebabkan ketidakseimbangan dalam performa model.)
5. Akurasi: Secara keseluruhan, model memiliki akurasi 95%, yang menunjukkan bahwa mayoritas prediksi model adalah benar.
6. Macro Avg: Rata-rata dari precision, recall, dan F1-score tanpa mempertimbangkan proporsi kelas.
 - Precision: 0.95.
 - Recall: 0.87 (lebih rendah karena recall untuk churn lebih rendah).
 - F1-score: 0.90.
7. Weighted Avg: Rata-rata dari precision, recall, dan F1-score, dengan mempertimbangkan proporsi kelas.
 - Weighted Avg Recall: 0.95, karena Existing Customer (0) mendominasi jumlah data.

Confusion Matrix

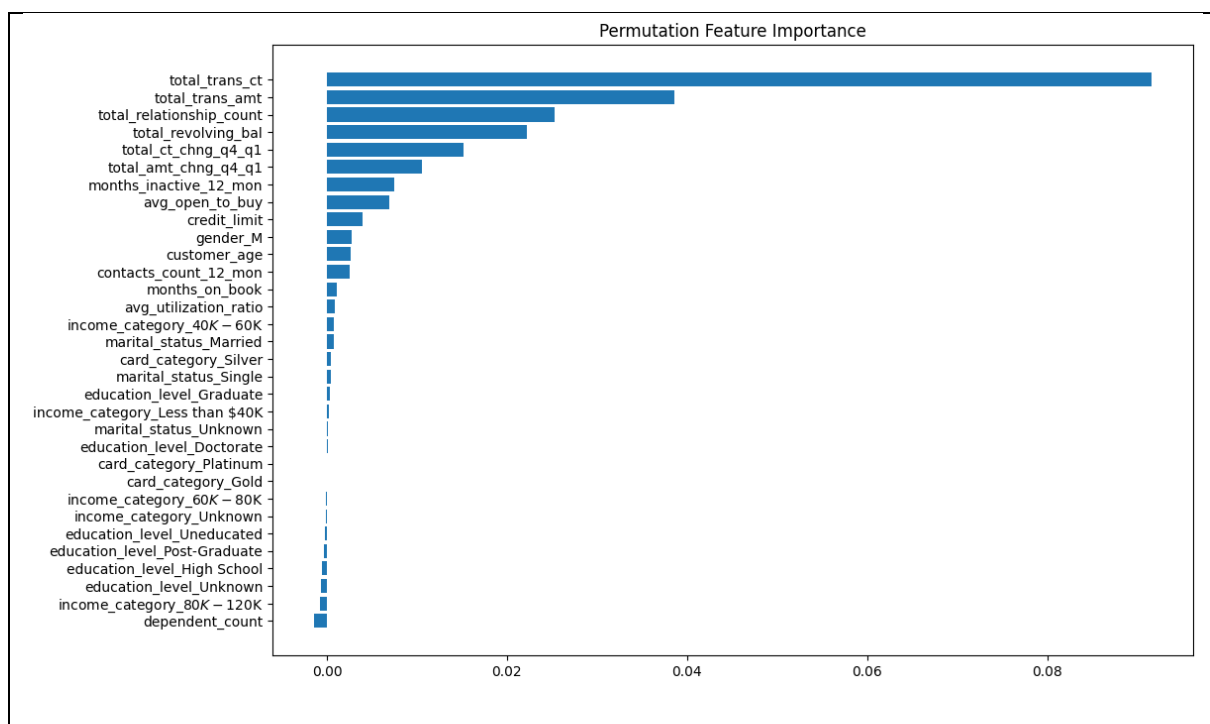
Confusion matrix memberikan detail jumlah prediksi benar dan salah untuk kedua kelas.

Penjelasan Tiap Komponen

1. True Negatives (TN) = 1682
 - Pelanggan yang benar-benar Existing Customer (tidak churn) dan diprediksi sebagai Existing Customer.
 - Model memiliki kemampuan yang sangat baik dalam mengidentifikasi pelanggan yang tidak churn.
2. False Positives (FP) = 17

- Pelanggan yang sebenarnya Existing Customer (tidak churn) tetapi diprediksi sebagai Attrited Customer (churn).
 - Angka ini sangat kecil, menunjukkan bahwa model sangat jarang memberikan peringatan palsu bahwa pelanggan akan churn.
3. False Negatives (FN) = 80
- Pelanggan yang sebenarnya Attrited Customer (churn) tetapi diprediksi sebagai Existing Customer (tidak churn).
 - Angka ini cukup besar, menunjukkan bahwa model masih gagal mendeteksi beberapa pelanggan yang berisiko churn. Ini adalah area yang perlu ditingkatkan, terutama untuk recall pada kelas churn.
4. True Positives (TP) = 247
- Pelanggan yang benar-benar Attrited Customer (churn) dan diprediksi dengan benar sebagai Attrited Customer.
 - Model mampu mengidentifikasi sebagian besar pelanggan churn, meskipun tidak semuanya (recall untuk churn hanya 76%).

5. Fitur Relevan



Analisa :

Permutation Feature Importance mengukur seberapa signifikan pengaruh fitur terhadap prediksi model secara keseluruhan. berikut adalah interpretasi dari fitur-fitur paling penting dalam prediksi churn:

Fitur-Fitur dengan Pengaruh Paling Besar

1. total_trans_ct (Jumlah Transaksi):
 - Interpretasi:

Pelanggan yang memiliki jumlah transaksi lebih rendah berpotensi lebih besar untuk meninggalkan platform. Frekuensi transaksi mencerminkan tingkat penggunaan layanan.

- Insight:
Aktivitas transaksi yang rendah dapat menjadi indikator bahwa pelanggan mulai kehilangan minat atau merasa kurang terlibat dengan layanan.
 - Langkah yang Bisa Dilakukan:
Tawarkan program loyalitas berbasis frekuensi transaksi, seperti cashback atau reward untuk setiap transaksi tertentu, untuk mendorong pelanggan lebih aktif.
2. total_trans_amt (Total Nominal Transaksi):
- Interpretasi:
Nilai transaksi yang rendah sering kali berhubungan dengan risiko churn. Pelanggan yang tidak memanfaatkan platform secara signifikan cenderung lebih mudah meninggalkan layanan.
 - Insight:
Pelanggan dengan transaksi kecil mungkin kurang melihat nilai dari layanan yang diberikan.
 - Langkah yang Bisa Dilakukan:
Berikan promosi berbasis volume transaksi, seperti diskon untuk transaksi dengan nilai tertentu, guna meningkatkan partisipasi finansial pelanggan.
3. total_relationship_count (Jumlah Hubungan Produk):
- Interpretasi:
Pelanggan dengan jumlah hubungan produk yang lebih sedikit (misalnya, hanya menggunakan 1-2 layanan) cenderung kurang terikat dengan platform.
 - Insight:
Semakin banyak produk yang digunakan pelanggan, semakin tinggi tingkat loyalitas mereka terhadap platform.
 - Langkah yang Bisa Dilakukan:
Promosikan kombinasi layanan melalui penawaran bundling atau diskon untuk pelanggan yang mencoba layanan baru.
4. total_revolving_bal (Saldo Revolving):
- Interpretasi:
Saldo revolving yang rendah menunjukkan pelanggan kurang memanfaatkan fasilitas kredit atau layanan terkait.
 - Insight:
Pelanggan yang jarang memanfaatkan layanan ini mungkin merasa tidak relevan dengan kebutuhan mereka.
 - Langkah yang Bisa Dilakukan:
Edukasi pelanggan tentang manfaat saldo revolving dan cara optimal memanfaatkan layanan.
5. total_ct_chng_q4_q1 (Perubahan Jumlah Transaksi antara Q4 dan Q1):

- Interpretasi:
Penurunan signifikan pada jumlah transaksi dari kuartal sebelumnya adalah sinyal awal bahwa pelanggan sedang mengurangi aktivitasnya.
 - Insight:
Tren ini sering kali merupakan indikator kuat bahwa pelanggan sedang menuju churn.
 - Langkah yang Bisa Dilakukan:
Lakukan kampanye re-engagement kepada pelanggan dengan pola penurunan aktivitas untuk menarik mereka kembali menggunakan platform.
6. total_amt_chng_q4_q1 (Perubahan Nominal Transaksi antara Q4 dan Q1):
- Interpretasi:
Penurunan nilai transaksi adalah tanda bahwa pelanggan mungkin kurang puas atau tidak lagi membutuhkan layanan secara intens.
 - Insight:
Sama seperti jumlah transaksi, nominal transaksi juga merupakan indikator penting penurunan minat pelanggan.
 - Langkah yang Bisa Dilakukan:
Tawarkan promosi khusus bagi pelanggan dengan penurunan nilai transaksi untuk mendorong mereka tetap aktif.

Fitur-Fitur dengan Pengaruh Menengah

7. months_inactive_12_mon (Bulan Tidak Aktif dalam 12 Bulan Terakhir):
- Interpretasi:
Pelanggan yang tidak aktif selama lebih dari 3 bulan menunjukkan risiko churn yang lebih tinggi.
 - Insight:
Periode tidak aktif yang lama sering kali mengindikasikan hilangnya minat atau kurangnya interaksi dengan platform.
 - Langkah yang Bisa Dilakukan:
Kirimkan pesan personalisasi atau insentif untuk mengaktifkan kembali pelanggan sebelum mereka memasuki fase tidak aktif yang lebih lama.
8. avg_open_to_buy (Kredit yang Tersedia):
- Interpretasi:
Pelanggan dengan kredit yang tersedia lebih tinggi cenderung lebih stabil, sementara pelanggan dengan kredit rendah mungkin lebih berisiko churn.
 - Insight:
Kredit yang tersedia adalah indikator kemampuan pelanggan untuk terus bertransaksi, terutama untuk produk keuangan.
 - Langkah yang Bisa Dilakukan:
Tingkatkan keterlibatan pelanggan dengan penawaran yang relevan dengan status kredit mereka.
9. credit_limit (Limit Kredit):

- Interpretasi:
Pelanggan dengan limit kredit lebih tinggi lebih jarang churn dibandingkan dengan pelanggan dengan limit kredit rendah.
 - Insight:
Limit kredit mencerminkan potensi finansial pelanggan, di mana pelanggan dengan limit rendah mungkin merasa kurang fleksibel.
 - Langkah yang Bisa Dilakukan:
Evaluasi pelanggan untuk meningkatkan limit kredit berdasarkan pola penggunaan mereka, yang dapat meningkatkan loyalitas.
10. gender_M (Jenis Kelamin Pria):
- Interpretasi:
Faktor gender memberikan kontribusi kecil, namun pria cenderung memiliki pola perilaku penggunaan yang berbeda.
 - Insight:
Preferensi layanan mungkin sedikit berbeda antara pelanggan pria dan wanita.
 - Langkah yang Bisa Dilakukan:
Kustomisasi kampanye pemasaran dengan mempertimbangkan pola penggunaan berdasarkan gender.
11. customer_age (Usia Pelanggan):
- Interpretasi:
Usia pelanggan memiliki pengaruh kecil, tetapi generasi yang lebih muda dan lebih tua mungkin memiliki perilaku penggunaan yang berbeda.
 - Insight:
Usia dapat memengaruhi kebutuhan pelanggan terhadap layanan tertentu.
 - Langkah yang Bisa Dilakukan:
Sesuaikan promosi dan komunikasi berdasarkan kelompok usia untuk menarik perhatian mereka.

Fitur-Fitur dengan Pengaruh Kecil

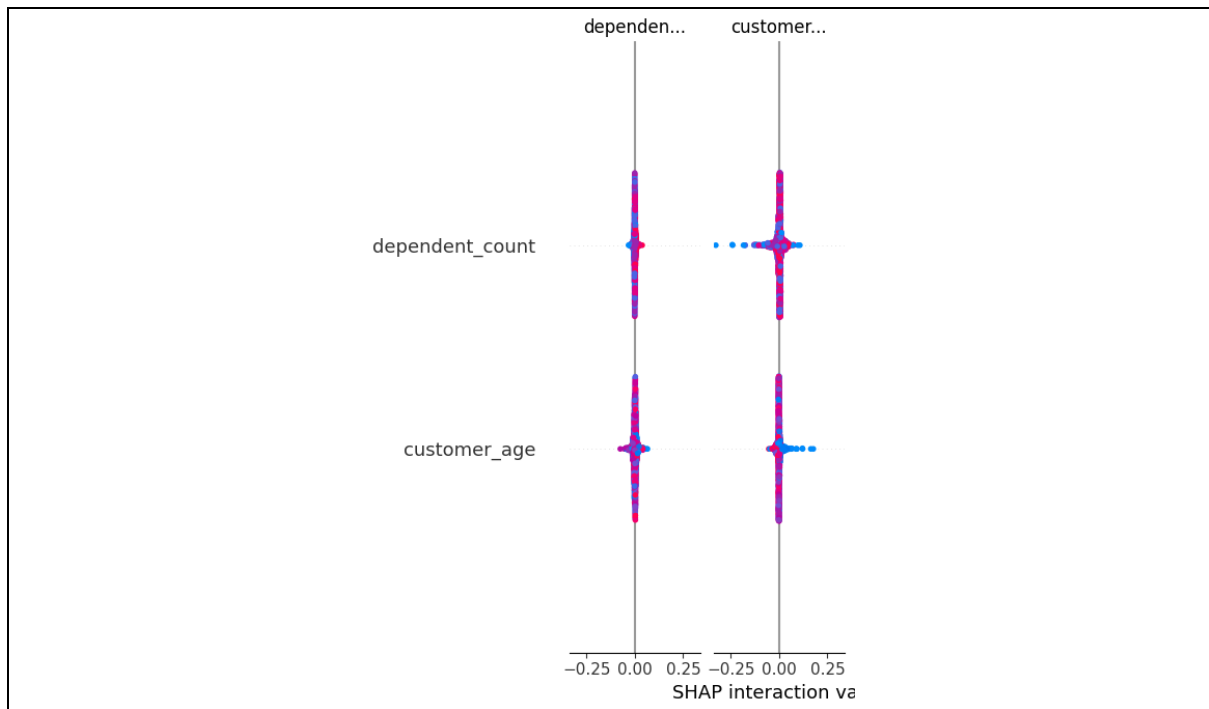
12. contacts_count_12_mon (Jumlah Interaksi dalam 12 Bulan Terakhir):
- Interpretasi:
Pelanggan dengan lebih sedikit interaksi menunjukkan kurangnya keterlibatan.
 - Insight:
Kurangnya komunikasi bisa menyebabkan pelanggan merasa diabaikan atau kurang relevan.
 - Langkah yang Bisa Dilakukan:
Tingkatkan komunikasi proaktif dengan pelanggan melalui email, telepon, atau notifikasi.
13. months_on_book (Jumlah Bulan Sebagai Pelanggan):
- Interpretasi:
Pelanggan baru sering kali lebih rentan untuk churn karena loyalitas mereka belum terbangun.

- Insight:
Membina hubungan dengan pelanggan baru sangat penting untuk mencegah churn dini.
- Langkah yang Bisa Dilakukan:
Perkuat program onboarding untuk memberikan pengalaman pertama yang berkesan bagi pelanggan baru.

14. avg_utilization_ratio (Rasio Pemanfaatan Kredit):

- Interpretasi:
Pelanggan dengan rasio pemanfaatan kredit sangat rendah atau sangat tinggi menunjukkan perilaku yang berbeda.
- Insight:
Rasio ini memberikan sinyal tentang tingkat aktivitas pelanggan terhadap layanan.
- Langkah yang Bisa Dilakukan:
Dorong pelanggan dengan rasio rendah untuk lebih memanfaatkan kredit mereka melalui edukasi atau insentif.

6. SHAP



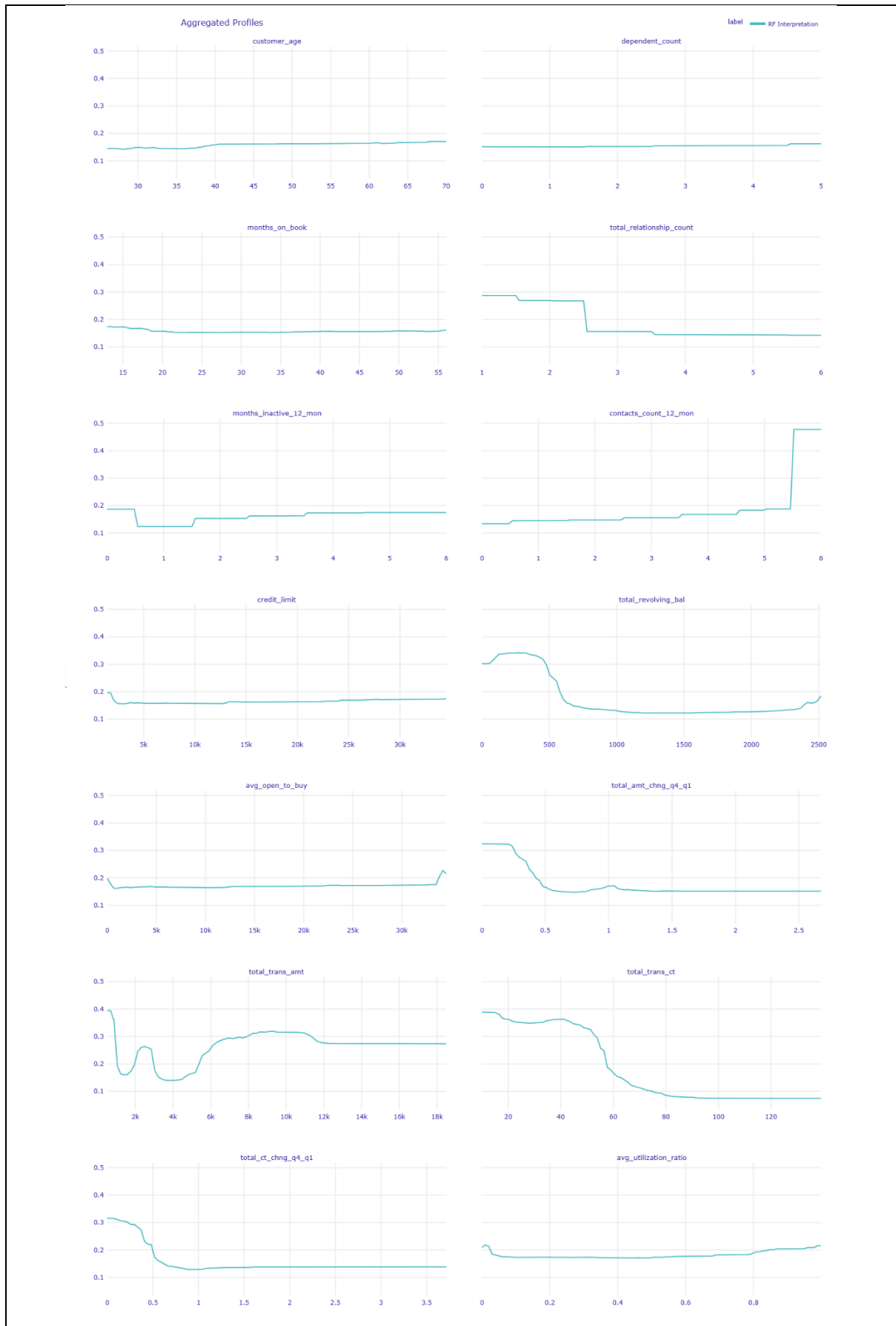
Analisa :

1. `dependent_count`:
 - Kontribusi rendah terhadap churn: Data menunjukkan bahwa fitur `dependent_count` memiliki interaksi yang sangat terbatas, dengan banyak titik yang terpusat di sekitar nilai SHAP dekat 0. Ini mengindikasikan bahwa jumlah tanggungan tidak terlalu mempengaruhi keputusan churn pada model ini.
 - Distribusi: Nilai-nilai SHAP untuk fitur ini terdistribusi secara merata dengan sedikit perbedaan antar kelas churn dan tidak churn.
2. `customer_age`:
 - Pengaruh yang lebih signifikan terhadap churn. Titik-titik pada plot ini menunjukkan perbedaan yang lebih besar antara nilai SHAP untuk churn dan non-churn. Namun, meskipun pengaruhnya lebih signifikan dibandingkan `dependent_count`, kontribusinya terhadap churn masih relatif kecil, yang tercermin dari nilai SHAP yang tersebar.
 - Pola Umum: Nilai SHAP untuk usia cenderung lebih tinggi untuk pelanggan yang lebih muda, tetapi pengaruhnya terhadap model tidak cukup besar untuk membuat keputusan prediksi churn yang kuat.

Kesimpulan :

- Berdasarkan hasil visualisasi SHAP ini, dapat disimpulkan bahwa `dependent_count` dan `customer_age` tidak terlalu berpengaruh terhadap prediksi churn jika dibandingkan dengan fitur lain seperti jumlah transaksi atau jumlah produk yang digunakan. Namun, meskipun kontribusi fitur-fitur ini lebih kecil, mereka tetap penting untuk analisis lanjutan, khususnya dalam konteks segmentasi nasabah atau pembuatan strategi pemasaran yang lebih personal.

7. PDP



Analisa :

3. total_trans_ct (Jumlah Transaksi)

- Hubungan dengan Probabilitas Churn:
 - Pelanggan dengan jumlah transaksi rendah (<40 transaksi per tahun) memiliki risiko churn yang tinggi.
 - Probabilitas churn menurun drastis saat jumlah transaksi mencapai 40–60 transaksi.
 - Di atas 60 transaksi, risiko churn hampir mendekati nol.
- Kesimpulan:
 - Jumlah transaksi adalah indikator utama keterlibatan pelanggan. Semakin tinggi jumlah transaksi, semakin kecil kemungkinan churn.
- Langkah yang Bisa Dilakukan:
 - Fokus pada pelanggan dengan transaksi rendah (kurang dari 40 transaksi) melalui insentif seperti cashback, diskon, atau hadiah loyalitas.

4. total_trans_amt (Total Nominal Transaksi)

- Hubungan dengan Probabilitas Churn:
 - Risiko churn tinggi untuk pelanggan dengan nominal transaksi rendah (<\$6.000 per tahun).
 - Risiko menurun signifikan ketika nominal transaksi berada di antara \$6.000–\$12.000.
 - Pelanggan dengan transaksi di atas \$15.000 memiliki risiko churn yang sangat rendah.
- Kesimpulan:
 - Nominal transaksi menunjukkan nilai dan aktivitas pelanggan di platform. Pelanggan dengan transaksi besar lebih loyal.
- Langkah yang Bisa Dilakukan:
 - Tawarkan program insentif berbasis volume transaksi untuk pelanggan dengan nominal rendah, seperti diskon untuk pembelian besar.

5. total_relationship_count (Jumlah Hubungan Produk)

- Hubungan dengan Probabilitas Churn:
 - Pelanggan dengan 1–2 hubungan produk memiliki risiko churn yang tinggi.
 - Probabilitas churn menurun tajam jika pelanggan memiliki 3 atau lebih hubungan produk.
 - Pelanggan dengan 5 atau lebih hubungan produk hampir tidak memiliki risiko churn.
- Kesimpulan:
 - Pelanggan dengan keterlibatan produk yang lebih banyak memiliki loyalitas yang lebih tinggi.
- Langkah yang Bisa Dilakukan:
 - Dorong pelanggan dengan hubungan produk rendah untuk mencoba layanan tambahan melalui bundling atau promosi khusus.

6. months_inactive_12_mon (Bulan Tidak Aktif dalam 12 Bulan Terakhir)

- Hubungan dengan Probabilitas Churn:

- Pelanggan yang tidak aktif selama lebih dari 2 bulan mulai menunjukkan risiko churn yang meningkat.
 - Risiko churn menjadi sangat tinggi bagi pelanggan yang tidak aktif selama 4–6 bulan.
 - Kesimpulan:
 - Ketidaktaktifan adalah indikator awal churn. Semakin lama tidak aktif, semakin besar risiko churn.
 - Langkah yang Bisa Dilakukan:
 - Lakukan kampanye re-engagement untuk pelanggan yang tidak aktif 2 bulan atau lebih.
 - Hubungi pelanggan secara langsung jika mereka telah tidak aktif selama 4 bulan untuk mengetahui alasannya.
7. total_ct_chng_q4_q1 (Perubahan Jumlah Transaksi antara Q4 dan Q1)
- Hubungan dengan Probabilitas Churn:
 - Penurunan jumlah transaksi antar kuartal adalah sinyal kuat risiko churn.
 - Peningkatan jumlah transaksi menunjukkan loyalitas pelanggan yang lebih tinggi.
 - Kesimpulan:
 - Perubahan negatif pada aktivitas transaksi mencerminkan penurunan keterlibatan pelanggan.
 - Langkah yang Bisa Dilakukan:
 - Monitor pelanggan dengan tren penurunan transaksi antar kuartal.
 - Tawarkan insentif untuk memotivasi mereka kembali aktif bertransaksi.
8. total_amt_chng_q4_q1 (Perubahan Nominal Transaksi antara Q4 dan Q1)
- Hubungan dengan Probabilitas Churn:
 - Penurunan nominal transaksi antar kuartal meningkatkan risiko churn.
 - Peningkatan nominal transaksi menunjukkan keterlibatan pelanggan yang lebih tinggi.
 - Kesimpulan:
 - Penurunan nilai transaksi adalah tanda awal ketidakpuasan atau kehilangan minat pelanggan.
 - Langkah yang Bisa Dilakukan:
 - Lakukan kampanye khusus untuk pelanggan dengan tren penurunan nominal transaksi, seperti promosi berbasis volume transaksi.
9. avg_utilization_ratio (Rasio Pemanfaatan Kredit)
- Hubungan dengan Probabilitas Churn:
 - Pelanggan dengan rasio pemanfaatan kredit yang sangat rendah (<10%) atau sangat tinggi (>90%) memiliki risiko churn yang lebih tinggi.
 - Pelanggan dengan rasio pemanfaatan moderat (30–60%) memiliki risiko churn yang rendah.
 - Kesimpulan:
 - Rasio pemanfaatan kredit yang ekstrem mencerminkan pelanggan yang kurang aktif atau berada di bawah tekanan finansial.
 - Langkah yang Bisa Dilakukan:

- Dorong pelanggan dengan rasio rendah untuk lebih sering menggunakan kredit melalui promosi.
 - Berikan solusi fleksibilitas pembayaran untuk pelanggan dengan rasio tinggi.
10. months_on_book (Jumlah Bulan Sebagai Pelanggan)
- Hubungan dengan Probabilitas Churn:
 - Pelanggan baru (<12 bulan) memiliki risiko churn yang lebih tinggi dibandingkan pelanggan lama.
 - Risiko churn menurun seiring bertambahnya waktu pelanggan menggunakan platform.
 - Kesimpulan:
 - Pelanggan baru membutuhkan perhatian ekstra untuk membangun loyalitas mereka.
 - Langkah yang Bisa Dilakukan:
 - Perkuat program onboarding untuk pelanggan baru.
 - Berikan insentif loyalitas untuk membantu pelanggan melewati masa kritis awal mereka.
11. contacts_count_12_mon (Jumlah Interaksi dalam 12 Bulan)
- Hubungan dengan Probabilitas Churn:
 - Pelanggan dengan interaksi rendah (<3 kontak per tahun) memiliki risiko churn yang lebih tinggi.
 - Pelanggan dengan lebih banyak interaksi menunjukkan loyalitas yang lebih tinggi.
 - Kesimpulan:
 - Frekuensi interaksi mencerminkan tingkat perhatian yang diberikan kepada pelanggan.
 - Langkah yang Bisa Dilakukan:
 - Tingkatkan komunikasi dengan pelanggan melalui kampanye email, notifikasi, atau layanan pelanggan proaktif.
12. credit_limit (Limit Kredit)
- Hubungan dengan Probabilitas Churn:
 - Pelanggan dengan limit kredit lebih rendah memiliki risiko churn yang lebih tinggi.
 - Pelanggan dengan limit kredit lebih tinggi cenderung lebih loyal.
 - Kesimpulan:
 - Limit kredit yang rendah dapat menjadi penghambat bagi pelanggan untuk menggunakan layanan secara optimal.
 - Langkah yang Bisa Dilakukan:
 - Evaluasi pelanggan untuk meningkatkan limit kredit mereka jika memenuhi kriteria.
13. avg_open_to_buy (Kredit yang Tersedia)
- Hubungan dengan Probabilitas Churn:
 - Pelanggan dengan kredit tersedia rendah memiliki risiko churn yang lebih tinggi.
 - Pelanggan dengan kredit tersedia tinggi menunjukkan loyalitas yang lebih besar.
 - Kesimpulan:
 - Kredit yang tersedia mencerminkan kemampuan finansial pelanggan untuk terus aktif menggunakan layanan.

- Langkah yang Bisa Dilakukan:
 - Tawarkan promosi berbasis kredit untuk pelanggan dengan kredit tersedia rendah.

14. dependent_count (Jumlah Tanggungan)

- Hubungan dengan Probabilitas Churn:
 - Jumlah tanggungan memiliki pengaruh yang kecil terhadap churn.
 - Pelanggan dengan tanggungan lebih banyak cenderung memiliki pengeluaran yang lebih stabil.
- Kesimpulan:
 - Meskipun pengaruhnya kecil, jumlah tanggungan dapat membantu dalam segmentasi pelanggan.
- Langkah yang Bisa Dilakukan:
 - Segmentasikan pelanggan berdasarkan jumlah tanggungan untuk mencocokkan promosi yang relevan.

15. education_level (Tingkat Pendidikan)

- Hubungan dengan Probabilitas Churn:
 - Pelanggan dengan pendidikan lebih tinggi cenderung lebih loyal.
 - Pengaruh pendidikan terhadap churn relatif kecil tetapi dapat digunakan untuk segmentasi.
- Kesimpulan:
 - Tingkat pendidikan membantu memahami preferensi pelanggan.
- Langkah yang Bisa Dilakukan:
 - Gunakan tingkat pendidikan untuk merancang komunikasi pemasaran yang lebih relevan.

16. income_category (Kategori Pendapatan)

- Hubungan dengan Probabilitas Churn:
 - Pelanggan dengan pendapatan lebih tinggi menunjukkan risiko churn yang lebih rendah.
 - Pelanggan dengan pendapatan rendah cenderung lebih rentan terhadap churn.
- Kesimpulan:
 - Pendapatan adalah faktor pendukung dalam memahami pola perilaku pelanggan.
- Langkah yang Bisa Dilakukan:
 - Tawarkan promosi yang sesuai untuk pelanggan dengan pendapatan rendah.

KESIMPULAN

1. Karakteristik Nasabah yang Rentan Churn

- **Distribusi Usia:** Nasabah usia menengah (40–50 tahun) memiliki jumlah churn tertinggi. Kelompok usia 50–60 tahun juga menunjukkan risiko churn yang cukup besar, sedangkan usia lebih tua (60–80 tahun) cenderung lebih stabil.
 - Analisis: Kelompok usia menengah mungkin memiliki ekspektasi yang lebih tinggi terhadap layanan bank, sehingga lebih rentan churn jika ekspektasi tidak terpenuhi.
 - Tindakan: Fokus pada program peningkatan kepuasan untuk kelompok usia ini, seperti layanan personalisasi atau penawaran produk sesuai kebutuhan mereka.
- **Tingkat Pendidikan:** Tingkat pendidikan yang lebih tinggi, seperti "Doctorate" dan "Post-Graduate," menunjukkan proporsi churn yang lebih besar meskipun jumlah nasabahnya kecil.
 - Analisis: Nasabah berpendidikan tinggi cenderung memiliki ekspektasi layanan yang tinggi, yang dapat meningkatkan risiko churn jika layanan tidak sesuai ekspektasi.
 - Tindakan: Pastikan kualitas layanan bank memenuhi kebutuhan nasabah dengan pendidikan tinggi melalui pendekatan berbasis konsultasi atau layanan premium.
- **Pendapatan:** Kategori pendapatan rendah (misalnya, di bawah \$40K) memiliki jumlah churn tertinggi.
 - Analisis: Nasabah berpendapatan rendah mungkin lebih sensitif terhadap biaya layanan atau produk yang tidak memberikan nilai tambah.
 - Tindakan: Rancang program loyalitas atau produk yang terjangkau untuk menarik dan mempertahankan kelompok ini.
- **Gender:** Perempuan cenderung lebih sering churn dibandingkan laki-laki.
 - Analisis: Hal ini mungkin disebabkan oleh perbedaan preferensi layanan atau faktor lain seperti waktu dan fleksibilitas layanan.
 - Tindakan: Kustomisasi program untuk nasabah perempuan dengan mempertimbangkan preferensi mereka, seperti layanan yang lebih mudah diakses atau berbasis digital.

2. Insight dari Pola Churn

- **Durasi Keanggotaan:** Nasabah yang lebih lama menjadi pelanggan tidak selalu memiliki risiko churn yang lebih kecil. Ada fluktuasi churn berdasarkan bulan tertentu.
 - Analisis: Faktor seperti pengalaman buruk pada layanan atau kebutuhan yang berubah dapat memengaruhi churn meskipun durasi keanggotaan panjang.
 - Tindakan: Monitor pengalaman nasabah secara berkala, terutama pada bulan-bulan kritis di mana churn cenderung meningkat, dan berikan perhatian tambahan melalui komunikasi proaktif.
- **Frekuensi Interaksi:** Meningkatnya interaksi tidak selalu menurunkan risiko churn. Bahkan, beberapa kelompok dengan banyak interaksi menunjukkan peningkatan churn.
 - Analisis: Frekuensi interaksi tinggi mungkin terjadi karena keluhan atau masalah, yang dapat meningkatkan risiko churn jika tidak ditangani dengan baik.
 - Tindakan: Pastikan kualitas interaksi tetap tinggi, terutama untuk kelompok nasabah yang sering berhubungan dengan bank, seperti menangani keluhan dengan cepat dan profesional.

3. Performa Model Prediksi

- **Kinerja Model:** Model Random Forest memberikan akurasi sebesar 95%. Namun, recall untuk kelas churn (76%) lebih rendah dibandingkan kelas tidak churn (99%).
 - Analisis: Ketidakseimbangan jumlah data antara nasabah churn dan tidak churn memengaruhi performa model dalam mendeteksi churn.

- Tindakan: Gunakan teknik oversampling seperti SMOTE atau beri bobot lebih pada kelas churn untuk meningkatkan recall.
- Confusion Matrix:
 - False Negatives (FN): Ada 80 nasabah yang churn tetapi salah diklasifikasikan sebagai tidak churn.
 - True Positives (TP): Sebanyak 247 nasabah churn berhasil terdeteksi.
 - False Positives (FP): Hanya 17 nasabah yang salah diprediksi sebagai churn, menunjukkan model cukup akurat dalam menghindari peringatan palsu.
 - Tindakan: Fokus pada pengurangan FN untuk meningkatkan recall pada kelas churn melalui penyesuaian threshold prediksi.

4. Fitur-Fitur yang Paling Berpengaruh

- Fitur Utama:
 - total_trans_ct: Jumlah transaksi adalah indikator utama churn. Pelanggan dengan jumlah transaksi rendah (<40 transaksi/tahun) memiliki risiko churn tinggi.
 - total_trans_amt: Nilai nominal transaksi juga sangat penting. Nominal rendah (<\$6.000/tahun) meningkatkan risiko churn.
 - total_relationship_count: Jumlah hubungan produk berbanding terbalik dengan churn. Nasabah dengan lebih banyak hubungan produk (≥5 layanan) lebih kecil kemungkinan churn.

Langkah Strategis:

- Identifikasi pelanggan dengan jumlah transaksi rendah dan tawarkan insentif untuk meningkatkan aktivitas mereka.
- Promosikan layanan tambahan untuk meningkatkan jumlah hubungan produk.
- Fitur Menengah:
 - months_inactive_12_mon: Ketidaktifan selama lebih dari 3 bulan adalah tanda awal risiko churn.
 - avg_open_to_buy: Kredit yang tersedia tinggi menunjukkan nasabah yang stabil, sedangkan rendah meningkatkan risiko churn.

Langkah Strategis:

- Kirimkan notifikasi atau insentif re-engagement kepada nasabah tidak aktif.
- Berikan edukasi tentang manfaat pemanfaatan kredit untuk nasabah dengan kredit tersedia rendah.
- Fitur Minor:
 - dependent_count, education_level, dan income_category memiliki pengaruh kecil tetapi dapat digunakan untuk segmentasi tambahan.

Langkah Strategis: Gunakan fitur ini untuk membuat kampanye pemasaran yang lebih personal dan sesuai dengan kebutuhan segmen tertentu.

5. Rekomendasi Strategis

- Tingkatkan Loyalitas:
 - Berikan insentif kepada nasabah dengan risiko churn tinggi, seperti cashback, diskon, atau hadiah loyalitas.
 - Dorong penggunaan lebih banyak produk bank melalui bundling atau layanan khusus.
- Tangani Keluhan secara Efektif:
 - Pastikan tim layanan pelanggan merespons keluhan dengan cepat, terutama untuk kelompok nasabah yang sering berinteraksi dengan bank.

- Personalisasi Layanan:
 - Rancang program khusus untuk segmen nasabah tertentu, seperti perempuan, kelompok usia 40–50 tahun, atau mereka dengan pendapatan rendah.
- Fokus pada Nasabah Tidak Aktif:
 - Lakukan kampanye re-engagement secara rutin untuk nasabah yang tidak aktif selama lebih dari 2 bulan.