

Analisis Customer Churn

1. Dataset

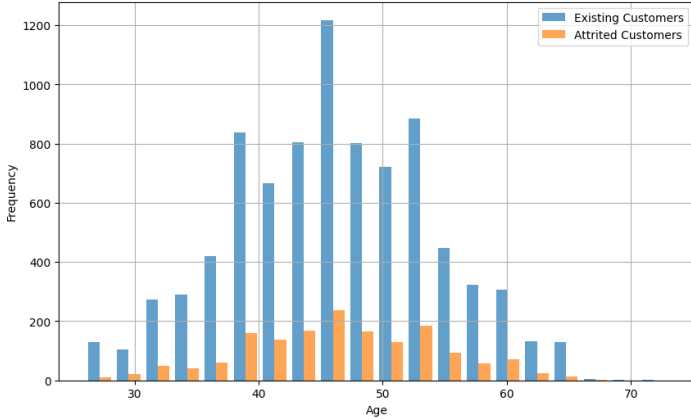
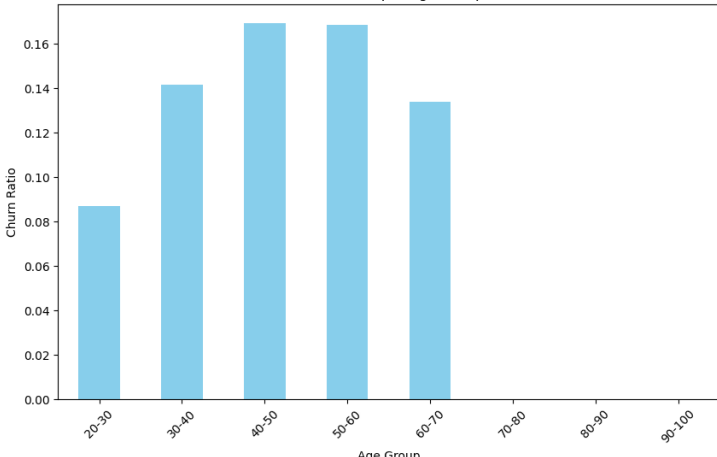
Dataset yang digunakan memiliki fitur-fitur sebagai berikut :

- **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.

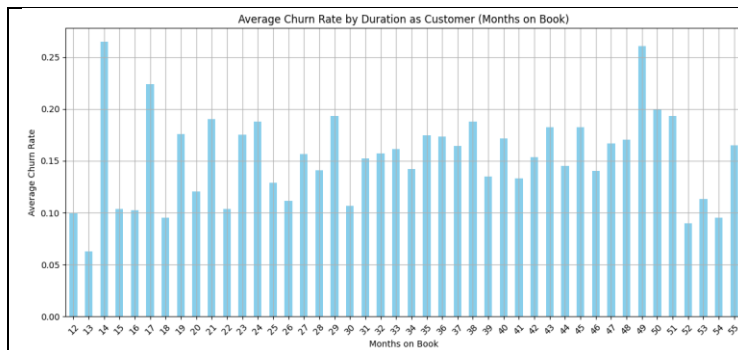
2. Pertanyaan Bisnis

- 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?

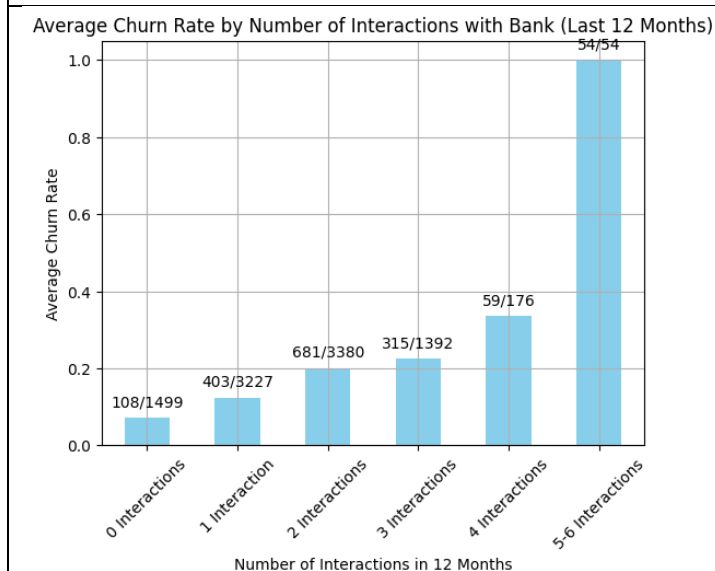
3. Visualisasi

Visualisasi	Penjelasan																																																								
<p>Distribution of Age by Customer Attrition Status</p>  <p>The bar chart displays the frequency of existing customers (blue bars) and attrited customers (orange bars) 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 in the 40-50 age range, while 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 displays the churn ratio per age group. The x-axis represents age groups from 20-30 to 90-100, and the y-axis represents the churn ratio from 0.00 to 0.16. 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 data-bbox="210 1292 825 1639"><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> <p>The contingency table shows the relationship between education level and income category. The color scale indicates the frequency of each combination, with red representing the highest frequency (1139 for Graduate/Less than \$40K) and blue representing the lowest frequency (30 for Post-Graduate/\$120K+).</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	<p>Beerdasarkan 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>900</td></tr></tbody></table>	Gender	attrition_flag = 0 (Count)	attrition_flag = 1 (Count)	M	4050	700	F	4400	900	<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	900																							



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.

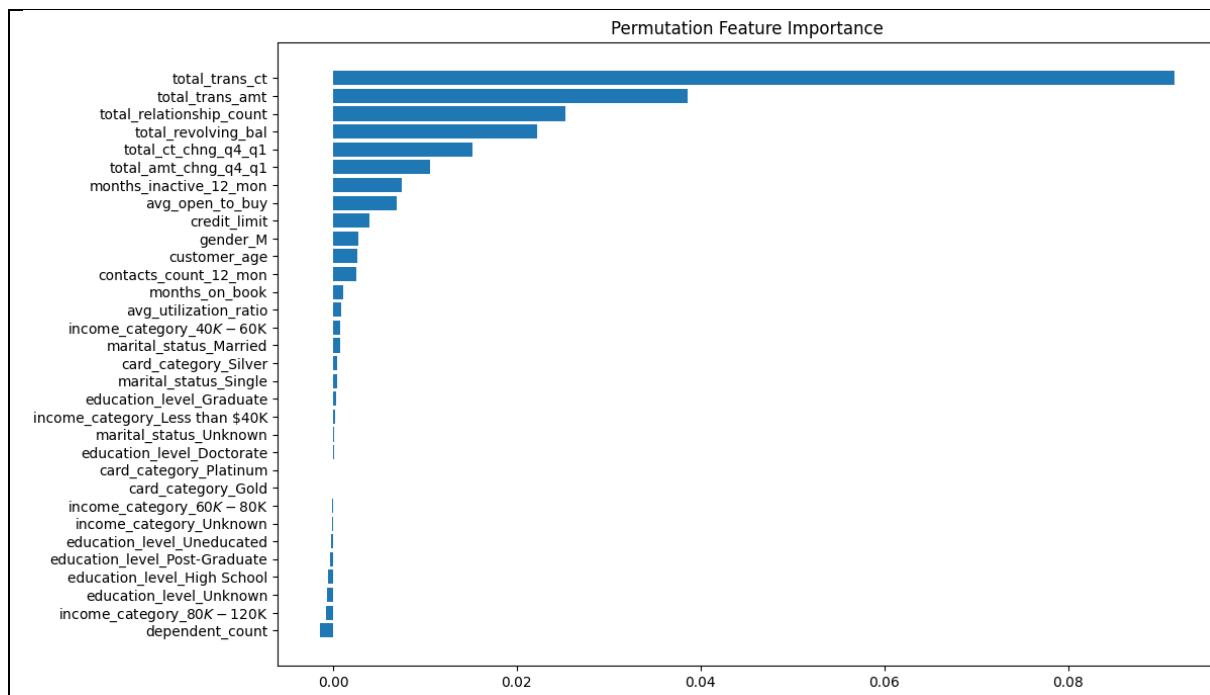
4. Performa Model

Model yang digunakan : RandomForestClassifier

Hasil Evaluasi :



5. Fitur Relevan



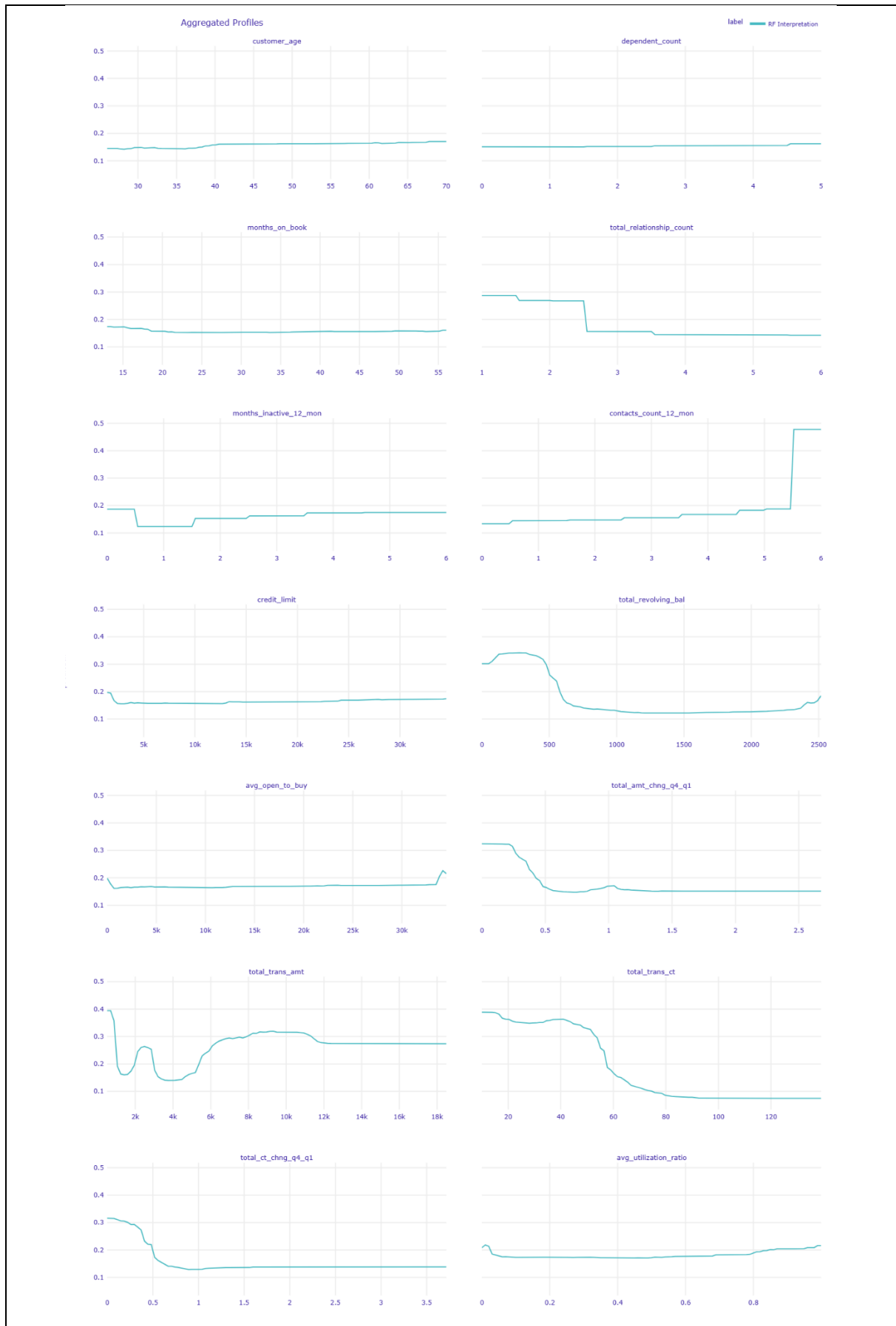
Analisa :

Berdasarkan hasil Permutation Feature Importance, fitur-fitur utama yang paling berpengaruh terhadap prediksi churn adalah:

1. total_trans_ct: Jumlah total transaksi.
2. total_trans_amt: Jumlah total nominal transaksi.
3. total_relationship_count: Jumlah total hubungan atau interaksi customer dengan platform.
4. total_revolving_bal: Total saldo revolving (utang berulang).
5. total_ct_chng_q4_q1: Perubahan jumlah transaksi dari kuartal 4 ke kuartal 1.
6. total_amt_chng_q4_q1: Perubahan nominal transaksi dari kuartal 4 ke kuartal 1.

Hasil ini menunjukkan bahwa metrik terkait frekuensi dan jumlah transaksi memiliki dampak terbesar terhadap prediksi churn.

6. PDP



Analisa :

1. total_trans_ct (Jumlah Transaksi)

- Probabilitas churn menurun tajam ketika jumlah transaksi meningkat, terutama hingga sekitar 60 transaksi.
- Setelah melewati angka ini, dampaknya terhadap probabilitas churn relatif stabil.

Rekomendasi:

- Fokus pada customer dengan jumlah transaksi rendah (< 60) untuk meningkatkan aktivitas mereka dengan promosi atau program loyalitas.

2. total_trans_amt (Jumlah Nominal Transaksi)

- Probabilitas churn menurun dengan meningkatnya jumlah nominal transaksi hingga sekitar \$10,000.
- Setelah titik ini, dampaknya cenderung konstan.

Rekomendasi:

- Memberikan insentif berbasis transaksi nominal untuk mendorong customer dengan nilai transaksi rendah.

3. total_relationship_count (Jumlah Hubungan)

- Probabilitas churn menurun secara signifikan ketika jumlah hubungan meningkat hingga sekitar 3 hubungan.
- Setelah itu, hubungan menjadi stabil.

Rekomendasi:

- Tingkatkan penggunaan produk lain di platform untuk menambah jumlah hubungan customer dengan layanan Anda.

4. total_ct_chng_q4_q1 dan total_amt_chng_q4_q1 (Perubahan Transaksi)

- Penurunan signifikan pada perubahan jumlah transaksi atau nominal transaksi dari Q4 ke Q1 meningkatkan risiko churn.
- Probabilitas churn menurun jika perubahan tetap atau meningkat.

Rekomendasi:

- Pantau pola perubahan transaksi dari waktu ke waktu. Hubungi customer dengan penurunan transaksi untuk mengidentifikasi masalah dan memberikan solusi proaktif.

5. months_inactive_12_mon (Bulan Tidak Aktif)

- Probabilitas churn meningkat seiring bertambahnya jumlah bulan tidak aktif, terutama setelah 2 bulan.

Rekomendasi:

- Buat kampanye re-engagement setelah customer tidak aktif selama 1-2 bulan untuk mencegah churn lebih lanjut.

6. total_revolving_bal (Saldo Revolving)

- Probabilitas churn meningkat jika saldo revolving meningkat, tetapi cenderung stabil setelah mencapai tingkat tertentu (1500).

Rekomendasi:

- Perkenalkan opsi pembayaran fleksibel atau insentif untuk membayar saldo revolving.

7. Interpretasi Model

Hasil evaluasi :

- Akurasi Tinggi: Model memiliki akurasi keseluruhan sebesar 95%, yang berarti sebagian besar prediksi model benar.
- Kemampuan Mendeteksi Customer yang Tidak Churn: Model sangat andal dalam mendeteksi customer yang masih aktif menggunakan platform (Existing Customers), dengan tingkat keberhasilan 99%.
- Kemampuan Mendeteksi Customer yang Berpotensi Churn: Model berhasil mendeteksi 76% customer yang benar-benar churn. Namun, sekitar 24% customer churn terlewatkan, yang berpotensi menyebabkan kehilangan peluang intervensi.
- Kesalahan Prediksi yang Minim: Hanya 17 customer aktif yang salah diprediksi akan churn, sehingga intervensi yang tidak perlu terhadap mereka sangat kecil.
- Faktor yang Perlu Diperhatikan: Customer dengan jumlah transaksi dan nominal transaksi yang rendah, atau yang menunjukkan pola penurunan transaksi, memiliki risiko churn lebih tinggi.

Feature Importance :

- Fitur Paling Berpengaruh:
 - 1) Jumlah Transaksi (total_trans_ct) adalah faktor terpenting yang menentukan apakah customer akan churn atau tetap menggunakan platform.
 - 2) Jumlah Nominal Transaksi (total_trans_amt) juga merupakan indikator utama, mencerminkan aktivitas finansial customer.
 - 3) Jumlah Hubungan dengan Platform (total_relationship_count), seperti jumlah produk yang digunakan, berperan penting dalam mempertahankan customer.
- Fitur Lain yang Relevan:
 - 1) Perubahan Transaksi Kuartalan (total_ct_chng_q4_q1 dan total_amt_chng_q4_q1): Customer yang mengalami penurunan transaksi dari kuartal sebelumnya lebih berisiko churn.
 - 2) Jumlah Bulan Tidak Aktif (months_inactive_12_mon): Customer yang tidak aktif selama beberapa bulan cenderung berhenti menggunakan platform.

Partial Dependence Plot (PDP) :

- Jumlah Transaksi Rendah Berisiko Tinggi: Customer dengan jumlah transaksi rendah (<60 transaksi) memiliki risiko churn yang jauh lebih tinggi dibandingkan mereka yang melakukan lebih banyak transaksi.
- Nominal Transaksi Rendah Berisiko Tinggi: Customer dengan nominal transaksi rendah (<\$10,000) lebih cenderung churn. Aktivitas finansial yang rendah menjadi sinyal utama risiko churn.
- Hubungan dengan Platform Sangat Penting: Customer dengan lebih banyak hubungan dengan platform (menggunakan lebih dari 3 layanan) memiliki risiko churn yang jauh lebih rendah.
- Penurunan Transaksi Sebagai Sinyal Peringatan: Penurunan signifikan dalam frekuensi atau nominal transaksi dari kuartal sebelumnya menunjukkan risiko churn yang lebih tinggi.
- Tidak Aktif Selama Beberapa Bulan: Customer yang tidak aktif selama lebih dari 2 bulan berturut-turut cenderung churn. Ini menunjukkan perlunya strategi re-engagement di awal ketidakaktifan.