**Analisis Customer Churn**

1. Dataset

Dataset yang digunakan memiliki fitut-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.

1. Pertanyaan Bisnis
2. Seperti apa distribusi usia para nasabah? Usia manakah yang cenderung memiliki kemungkinan churn terbesar?
3. Apakah terdapat korelasi antara tingkat pendidikan dan income category? Bagaimana korelasi dari kedua prediktor tersebut terhadap churn?
4. Apakah terdapat perbedaan yang signifikan antara gender dengan kemungkinan nasabah untuk churn?
5. Apakah semakin lama customer menjadi nasabah maka semakin kecil kemungkinan nasabah tersebut untuk churn?
6. Apakah semakin sering customer berinteraksi dengan bank dalam 1 tahun terakhir mempengaruhi besar atau kecilnya kemungkinan nasabah untuk churn?
7. Visualisasi

|  |  |
| --- | --- |
| Visualisai | Penjelasan |
|  | 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. |
|  | 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. |
|  | 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. |
|  | 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. |
|  | 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. |
|  | 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. |
|  | 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. |

1. Performa Model

Model yang digunakan : RandomForestClassifier

Hasil Evaluasi :

|  |  |
| --- | --- |
|  |  |

1. 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.

1. 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.

1. 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.

1. 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.

1. 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.

1. months\_inactive\_12\_mon (Bulan Tidak Aktif)

* Probabilitas churn meningkat seiring bertambahnya jumlah bulan tidak aktif, terutama setelah 3 bulan.

Rekomendasi:

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

1. 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.

1. 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 3 bulan berturut-turut cenderung churn. Ini menunjukkan perlunya strategi re-engagement di awal ketidakaktifan.