

Segmentasi Pelanggan dan Prediksi Churn: Strategi Optimalisasi Retensi Pelanggan dalam Industri Telekomunikasi

Mohamad Maulana Firdaus Ramadhan
Sekolah Teknik Elektro dan
Informatika
Institut Teknologi Bandung
Bandung, Indonesia
18222140@mahasiswa.itb.ac.id

Rizqi Andhika Pratama
Sekolah Teknik Elektro dan
Informatika
Institut Teknologi Bandung
Bandung, Indonesia
18222118@mahasiswa.itb.ac.id

Adinda Khairunnisa Indraputri
Sekolah Teknik Elektro dan
Informatika
Institut Teknologi Bandung
Bandung, Indonesia
18222104@mahasiswa.itb.ac.id

Abstract— Industri telekomunikasi menghadapi tantangan yang meningkat dalam mempertahankan pelanggan di tengah persaingan yang ketat dan perubahan preferensi pelanggan. Dalam konteks ini, studi ini dirancang untuk memberikan wawasan mendalam tentang perilaku pelanggan melalui segmentasi dan prediksi churn. Melalui teknik analisis kluster, pelanggan dikelompokkan ke dalam tiga kategori utama: "Pelanggan Elite", "Pelanggan Menengah", dan "Pelanggan Pemula". Setiap klaster memiliki karakteristik unik, dengan "Pelanggan Elite" yang cenderung memberikan kontribusi tinggi bagi pendapatan perusahaan namun memiliki potensi churn yang lebih tinggi. Di sisi lain, ada juga sejumlah pelanggan yang menghadapi keterbatasan akses internet, mempengaruhi pilihan dan kepuasan mereka terhadap layanan. Pada bagian prediksi churn, berbagai model machine learning dievaluasi dengan LightGBM menunjukkan performa terbaik, dimana akurasi, AUC-ROC, dan nilai F1-score, yaitu masing masing sebesar 0.7942, 0.7477 dan 0.7473. Meskipun demikian, penelitian ini menyoroti kebutuhan akan strategi retensi yang lebih tepat sasaran dan inovasi layanan untuk meningkatkan kepuasan dan loyalitas pelanggan. Selanjutnya, penelitian ini merekomendasikan analisis lebih lanjut untuk mengeksplorasi faktor-faktor yang mempengaruhi keputusan pelanggan untuk churn dan bagaimana strategi yang lebih personal dapat diimplementasikan untuk mengatasi masalah tersebut..

Keywords—Industri telekomunikasi, Segmentasi pelanggan, Prediksi churn, LightGBM, Model machine learning

I. PENDAHULUAN

A. Latar Belakang

Telekomunikasi merupakan bidang yang sangat penting dalam kehidupan manusia. Bidang ini melibatkan pengiriman informasi, suara, dan data antara individu, perangkat, atau lokasi yang berbeda melalui berbagai media seperti kabel, gelombang radio, dan satelit. Bidang ini memiliki banyak peran penting dalam kehidupan manusia, seperti kemajuan teknologi, komunikasi jarak jauh, dan

pertumbuhan bisnis dan ekonomi. Karena itu, telekomunikasi telah menjadi elemen tak terpisahkan dalam kehidupan sehari-hari yang terus berkembang dan akan terus memiliki dampak signifikan dalam cara kita berinteraksi, bekerja, belajar, dan hidup.

Dalam era ekonomi pasar yang semakin kompetitif, pelanggan menjadi aset paling berharga bagi perkembangan perusahaan. Memahami preferensi berbagai kelompok konsumen, sikap belanja mereka, dan persepsi harga menjadi kunci kesuksesan pemasaran. Oleh karena itu, segmentasi pelanggan menjadi strategi penting bagi perusahaan untuk menemukan posisi pasar yang tepat bagi produk mereka.

Segmentasi pelanggan adalah metode untuk mengkategorikan pelanggan ke dalam kelompok-kelompok berdasarkan satu atau beberapa kriteria tertentu. Pelanggan dalam satu kelompok memiliki kesamaan karakteristik yang signifikan, sementara antar kelompok menunjukkan perbedaan yang jelas. Dengan segmentasi, perusahaan dapat menentukan kelompok pelanggan target yang paling potensial dan menciptakan peluang pemasaran yang lebih spesifik.

Tujuan dari segmentasi pelanggan antara lain:

1. Memahami komposisi keseluruhan dari pelanggan,
2. Mengenali karakteristik unik dari pelanggan berharga,
3. Mengetahui ciri khas dari pelanggan yang mungkin beralih,
4. Memahami pola konsumsi pelanggan,
5. Membedah karakteristik kelompok pelanggan berdasarkan tingkat kredit mereka.

Selain itu *Customer Segmentation* akan memberikan manfaat sebagai berikut

1. Meningkatkan Strategi Bisnis.

2. *User Experience* lebih baik.
3. Meningkatkan pendapatan
4. Tingkat Retensi yang Lebih Tinggi.
5. Biaya Akuisisi yang Lebih Rendah
6. Pemahaman Pelanggan yang Lebih Baik

Dalam era digital saat ini, banyak perusahaan yang mulai membangun basis data pelanggan mereka. Dengan teknologi dan algoritma machine learning, perusahaan dapat menganalisis data pelanggan dalam jumlah besar. Proses ini dikenal dengan istilah "data mining", yang memungkinkan perusahaan menemukan informasi berharga untuk menentukan posisi pasar dan membuat keputusan pemasaran yang tepat. Beberapa studi sebelumnya telah dilakukan dalam bidang ini, seperti manajemen hubungan pelanggan di sektor perbankan dan analisis pelanggan di industri kendaraan listrik.

Salah satu isu penting dalam industri telekomunikasi adalah fenomena "churn" atau pelanggan yang beralih dari satu penyedia layanan ke penyedia lain. Industri telekomunikasi menghadapi ancaman kehilangan hingga 25% pelanggannya setiap tahun. Fakta menunjukkan bahwa biaya akuisisi pelanggan baru bisa 5 hingga 10 kali lebih mahal daripada mempertahankan pelanggan lama.

Pada tahun 2017, perusahaan Kanada BCE dan Telus mengungkapkan bahwa biaya untuk mempertahankan seorang pelanggan mobile yang ada hampir 50 kali lebih murah dibandingkan dengan memperoleh pelanggan baru, dengan biaya retensi sebesar 11.04 dolar Kanada dan 11.74 dolar Kanada masing-masing, sementara rata-rata biaya perolehan pelanggan baru mencapai angka yang sangat tinggi, yaitu 521 dolar Kanada. Dalam pasar yang sudah jenuh, tampaknya akan lebih masuk akal bagi penyedia layanan untuk fokus pada mempertahankan pelanggan yang ada, daripada mencoba menarik pelanggan baru dari para pesaing.

Oleh karena itu, strategi pemasaran terbaik adalah mempertahankan pelanggan eksisting. Dalam hal ini, kecerdasan buatan (AI) memainkan peran penting dalam mengidentifikasi dan memprediksi perilaku churn pelanggan dengan efisien.

Selain itu, *Churn Prediction* juga memiliki beberapa manfaat lain seperti :

1. Mencegah *Customer Churn*.
2. Meningkatkan tingkat kepuasan pelanggan.
3. Playbook Layanan Pelanggan yang Lebih Baik.
4. Menciptakan peluang pertumbuhan.
5. Mengidentifikasi titik-titik masalah.

II. DATA AND METHODOLOGY

A. Dataset

Dalam rangka memahami perilaku serta preferensi pelanggan dalam industri telekomunikasi, kami mengumpulkan data dari pelanggan, sejumlah 7043 data pelanggan.

Dataset yang digunakan dalam penelitian ini berfokus pada fenomena "churn" di industri telekomunikasi, yakni kecenderungan pelanggan untuk meninggalkan suatu layanan. Setiap baris pada dataset merepresentasikan satu pelanggan, dengan tiap kolom menggambarkan atribut pelanggan tersebut. Adapun kolom "Churn Label" mengindikasikan data pelanggan yang telah meninggalkan layanan perusahaan sebulan sebelumnya.

Dataset kami mencakup informasi detail sebagai berikut:

- Customer ID: Identifikasi unik setiap pelanggan.
- Tenure Months: Durasi pelanggan berlangganan dengan perusahaan hingga akhir kuartal tertentu.
- Location: Lokasi tempat tinggal pelanggan.
- Device Class: Klasifikasi perangkat yang digunakan pelanggan.
- Produk (Games, Music, Education, Video): Menunjukkan apakah pelanggan menggunakan layanan internet untuk produk tertentu.
- Call Center: Apakah pelanggan memanfaatkan layanan call center.
- MyApp: Apakah pelanggan menggunakan layanan MyApp.
- Metode Pembayaran: Cara pelanggan membayar tagihannya.
- Monthly Purchase: Total belanja bulanan pelanggan untuk semua layanan dalam satuan ribuan IDR.
- Churn Label: Apakah pelanggan meninggalkan perusahaan dalam periode tertentu.
- Longitude & Latitude: Koordinat geografis tempat tinggal pelanggan.
- CLTV: Nilai seumur hidup pelanggan dalam satuan ribuan IDR, dihitung berdasarkan rumus perusahaan.

B. Metodologi

1) Customer Segmentation

a) Data Preparation

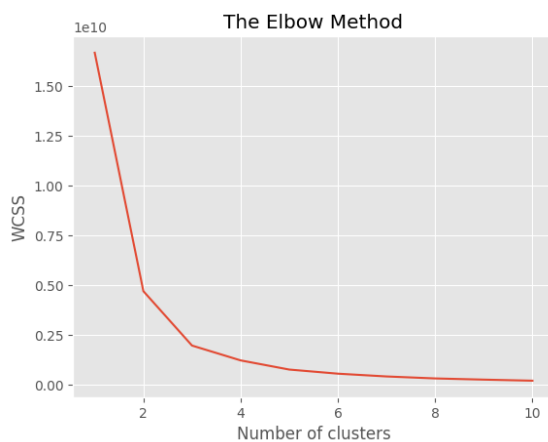
Sebagai langkah awal, penulis memilih beberapa kolom kategorikal untuk diaplikasikan one-hot encoding. Kolom-kolom ini termasuk 'Games Product', 'Device Class', 'Music Product', dan sebagainya. Metode one-hot encoding mengubah setiap kategori unik dalam suatu kolom menjadi kolom baru dan menggunakan format biner (0 atau 1) untuk menunjukkan kehadiran kategori tertentu pada baris tertentu.

Selanjutnya, kolom target 'Churn Label' diubah menjadi format numerik. Ini dilakukan dengan menggantikan 'No' dengan 0 dan 'Yes' dengan 1, menjadikannya lebih mudah untuk analisis dan pemodelan.

Akhirnya, kolom 'Customer ID' dihapus dari dataframe. karena tidak memberikan nilai prediktif untuk model dan dapat diabaikan dalam analisis lebih lanjut.

b) Modeling

Dalam mensegmentasi pelanggan, penulis menggunakan model KMeans Clustering yang merupakan salah satu metode klasterisasi yang populer dan efisien. Sebelum itu, penulis menggunakan metode Elbow Method untuk menentukan jumlah cluster yang optimal.



Gambar 1. Hasil Elbow Method

Berdasarkan grafik yang dihasilkan oleh metode "Elbow" diatas, tampak bahwa penurunan WCSS (Within-Cluster-Sum-of-Squares) mulai melambat dan mendekati titik stabil setelah tiga cluster. Oleh karena itu, dapat disimpulkan bahwa jumlah cluster yang optimal untuk dataset ini adalah tiga.

2) Churn Prediction

a) Data Preparation

Dalam tahap pemrosesan awal data pelanggan untuk *churn prediction*, penulis fokus pada ekstraksi informasi yang penting untuk prediksi "churn". Produk yang digunakan oleh pelanggan, seperti "Games Product", "Music Product", dan "Education Product", dihitung untuk menentukan jumlah produk yang digunakan. Selain itu, penulis juga memperhitungkan berbagai layanan lain yang digunakan oleh pelanggan, termasuk "Call Center" dan "Use MyApp". Untuk memahami hubungan antara lama berlangganan (tenure) dengan nilai pelanggan, penulis menciptakan fitur-fitur seperti "Is_Long_Tenure"

dan "Tenure_Category". Penulis juga mencoba memahami hubungan antara Lifetime Value (CLTV) pelanggan dengan pembelian bulanan melalui fitur-fitur seperti "CLTV_to_MonthlyPurchase_Ratio" dan "Tenure to CLTV Ratio".

Dengan adanya data yang mengindikasikan tidak ada layanan internet, seperti pada kolom "Games Product" dan "Music Product", kita memutuskan untuk mengkategorikannya sebagai 'No'. Untuk mempermudah pemodelan, kemudian penulis melakukan one-hot encoding pada kolom-kolom kategorikal. Kolom target, "Churn Label", diubah menjadi format biner untuk memudahkan prediksi.

Agar model lebih robust, fitur-fitur seperti "Monthly Purchase", "CLTV", dan rasio yang terkait dengan keduanya di-normalize menggunakan MinMax Scaler. Kemudian, data dipisahkan menjadi set pelatihan dan pengujian dengan rasio 80:20. Terakhir, untuk mengatasi data yang tidak seimbang, teknik SMOTE digunakan untuk menyeimbangkan kelas dalam data pelatihan. Ini akan memastikan bahwa model kita tidak bias terhadap kelas mayoritas saat pelatihan.

b) Modeling

Dalam penelitian kali ini, penulis menggunakan berbagai model machine learning untuk memprediksi "churn" pelanggan. Berikut adalah elaborasi formal dari setiap model:

Logistic Regression: Algoritma ini memfasilitasi prediksi biner dengan memodelkan probabilitas suatu sampel diklasifikasikan dalam suatu kategori spesifik.

Decision Tree: Model ini mengestimasi nilai variabel target dengan memformulasikan rangkaian aturan keputusan yang diperoleh dari fitur-fitur data.

Random Forest: Sebuah teknik ensemble yang terdiri dari multiplisitas pohon keputusan. Secara konvensional dilatih melalui metode "bootstrap aggregation", dimana setiap pohon diciptakan dari sampel bootstrap dari dataset pelatihan.

XGBoost: Merupakan kependekan dari "Extreme Gradient Boosting", algoritma ini adalah iterasi dari metode boosting gradien dengan fokus pada optimalisasi kecepatan dan performa.

LightGBM: Dikenal juga sebagai "Light Gradient Boosting Machine", merupakan framework boosting gradien yang mengimplementasikan algoritma berbasis pohon.

CatBoost: Spesialisasi dari algoritma boosting gradien yang dioptimalkan untuk fitur kategorikal,

menawarkan performa superior pada dataset dengan fitur kategorikal yang signifikan.

Gradient Boosting Machine: Mirip dengan XGBoost dan LightGBM, algoritma ini juga berlandaskan pada prinsip boosting gradien, tetapi dengan nuansa pendekatan yang berbeda.

AdaBoost: Atau "Adaptive Boosting", algoritma ini mengprioritaskan sampel yang sulit diprediksi dengan memberi bobot lebih selama fase pelatihan.

k-NN (k-Nearest Neighbors): Algoritma ini kategorisasi sampel berdasarkan karakteristik agragasi dari contoh-contoh terdekat dalam ruang fitur.

Support Vector Machine: Berupaya menemukan hyperplane yang optimal dalam memisahkan distribusi kelas data.

Naive Bayes: Sebuah ensemble algoritma yang berakar pada teorema Bayes dengan premis kemandirian antara fitur-fitur.

Ensemble Model dengan Hard Voting: Pendekatan ensemble ini mengagregasi prediksi dari berbagai model untuk menentukan prediksi akhir melalui mekanisme voting mayoritas.

Ensemble Model dengan Soft Voting: Analog dengan hard voting, namun probabilitas prediksi dari setiap model diaverage-kan untuk menghasilkan prediksi keseluruhan.

Model-model tersebut telah dikonfigurasi dengan set parameter spesifik dan siap dilatih dengan dataset pelatihan.

c) Evaluation

Dalam rangka evaluasi performa model, berbagai metrik telah diambil untuk menilai sejauh mana akurasi dan kapabilitas model dalam mengklasifikasikan data. Berikut adalah deskripsi formal dari prosedur tersebut:

Akurasi: Merupakan rasio antara jumlah prediksi yang benar dengan total prediksi yang dilakukan oleh model.

AUC (Area Under the Curve): Mengukur performa model dalam klasifikasi biner, khususnya sejauh mana model dapat membedakan antara dua kelas.

Classification Report: Adalah suatu metode untuk mendapatkan tiga metrik kunci lainnya, yaitu: precision, recall, dan F1-score.

- Precision (Presisi): Menunjukkan seberapa akurat model dalam memprediksi kelas positif. Diukur dengan rasio antara True Positive dengan jumlah dari True Positive dan False Positive.

- Recall (Sensitivitas): Mengukur proporsi dari seluruh sampel positif yang benar-benar diidentifikasi dengan benar.
- F1-Score: Adalah rata-rata harmonik dari precision dan recall, memberikan pandangan keseluruhan tentang keseimbangan antara kedua metrik tersebut.

Dalam konteks ini, kita mengambil rata-rata keseluruhan (macro average) untuk ketiga metrik tersebut.

III. Hasil

1) Customer Segmentation

Cluster	0	1	2
Tenure Months	33.653	18.399	41.058
Monthly Purchase (Thou. IDR)	83.934	79.173	88.0
CLTV (Predicted Thou. IDR)	5.638.742	3.624.698	7.287.048
Games Product_No	0.489	0.582	0.444
Games Product_No internet service	0.227	0.213	0.209
Games Product_Yes	0.284	0.205	0.347
Device Class_High End	0.433	0.435	0.449
Device Class_Low End	0.227	0.213	0.209
Device Class_Mid End	0.34	0.352	0.342
Music Product_No	0.424	0.536	0.384
Music Product_No internet service	0.227	0.213	0.209
Music Product_Yes	0.349	0.251	0.407
Payment Method_Credit	0.219	0.165	0.25
Payment Method_Debit	0.241	0.153	0.245
Payment Method_Digital Wallet	0.216	0.294	0.195
Payment Method_Pulsa	0.324	0.388	0.31
Education Product_No	0.423	0.523	0.396
Education Product_No internet service	0.227	0.213	0.209
Education Product_Yes	0.35	0.264	0.395
Use MyApp_No	0.374	0.478	0.357
Use MyApp_No internet service	0.227	0.213	0.209
Use MyApp_Yes	0.399	0.213	0.434
Video Product_No	0.38	0.476	0.362
Video Product_No internet service	0.227	0.213	0.209
Video Product_Yes	0.393	0.311	0.428
Call Center_No	0.704	0.785	0.663
Call Center_Yes	0.296	0.215	0.337
Longitude_106.816666	0.704	0.71	0.728
Longitude_107.60981	0.296	0.29	0.272
Latitude_-6.914744	0.296	0.29	0.272
Latitude_-6.2	0.704	0.71	0.728
Location_Bandung	0.296	0.29	0.272
Location_Jakarta	0.704	0.71	0.728
Churn Label_0	0.741	0.656	0.784
Churn Label_1	0.259	0.344	0.216

Gambar 2. Hasil Customer Segmentation

Gambar diatas merupakan hasil dari clustering yang telah dilakukan. Dimana terdapat 3 cluster berbeda 1 dengan yang lainnya dengan detail sebagai berikut beserta rekomendasi bisnis yang dapat diberikan untuk setiap cluster.

1. Cluster 0: "Pelanggan Menengah"

Cluster ini dikenal sebagai "Pelanggan Menengah" karena mereka berada di posisi tengah-tengah dalam hal masa berlangganan dan pembelian bulanan. Dengan masa berlangganan rata-rata sebesar 33,653 bulan, pelanggan ini menghabiskan rata-rata sekitar Rp. 83.934 setiap bulannya. Salah satu ciri khas dari cluster ini adalah kecenderungan mereka untuk tidak terlalu

sering menggunakan produk game, tetapi lebih sering menggunakan produk musik. Dari sisi metode pembayaran, pelanggan dalam cluster ini cenderung memilih metode pembayaran Pulsa. Selain itu, lokasi mayoritas pelanggan ini berada di Jakarta.

Rekomendasi Bisnis:

- Tawarkan program loyalitas untuk meningkatkan retensi dan engagement.
- Tingkatkan up-sell dengan menawarkan bundling produk game dan musik.
- Lakukan survei kepuasan pelanggan untuk mengetahui area yang bisa ditingkatkan.

2. Cluster 1: "Pelanggan Pemula"

Dikenal sebagai "Pelanggan Pemula", cluster ini terdiri dari pelanggan yang memiliki masa berlangganan paling pendek, namun menunjukkan aktivitas yang cukup tinggi dalam penggunaan beberapa produk. Dengan masa berlangganan rata-rata sebesar 18.399 bulan, mereka memiliki pembelian bulanan rata-rata sekitar Rp. 79.173. Pelanggan dalam cluster ini memiliki kecenderungan untuk lebih sering menggunakan produk video, namun tidak terlalu sering memanfaatkan produk musik. Sama seperti "Pelanggan Menengah", metode pembayaran yang paling sering digunakan adalah Pulsa, dan mayoritas dari mereka berlokasi di Jakarta.

Rekomendasi Bisnis:

- Buat program onboarding untuk meningkatkan awareness terhadap produk lain.
- Tawarkan promo khusus untuk produk video.
- Lakukan pendekatan pribadi melalui call center untuk meningkatkan retensi.
- Buatlah paket khusus untuk pelanggan baru dengan penawaran menarik di awal keanggotaan untuk meningkatkan retensi.
- Sebagai pelanggan yang mungkin belum terbiasa, tawarkan tutorial interaktif tentang produk dan layanan.
- Buat event atau webinar tentang tren digital terbaru untuk meningkatkan engagement dan edukasi.

3. Cluster 2: "Pelanggan Elite"

Cluster ini merupakan representasi dari "Pelanggan Elite" yang telah berlangganan dengan durasi paling lama dan cenderung memiliki pembelian bulanan yang lebih tinggi. Dengan masa berlangganan rata-rata mencapai 41.058 bulan, pelanggan dalam cluster ini menghabiskan rata-rata sekitar Rp. 88.000 setiap bulannya. Pelanggan dari cluster ini cenderung menggunakan beragam produk, termasuk game, musik, dan produk pendidikan. Seperti dua cluster lainnya, metode pembayaran yang paling sering dipilih adalah Pulsa, dan mayoritas dari pelanggan "elite" ini berlokasi di Jakarta.

Rekomendasi Bisnis:

- Pertahankan loyalitas mereka dengan memberikan penawaran eksklusif atau hadiah khusus.
- Berikan kesempatan untuk memberikan feedback atau saran untuk pengembangan produk.
- Pertimbangkan kerjasama dengan brand premium lain untuk memberikan penawaran khusus bagi segmen ini, meningkatkan persepsi eksklusivitas.
- Buatlah paket khusus untuk berbagai produk yang dipakai bisa berupa bundling atau promo tertentu.

Selain itu, perlu juga diperhatikan mengenai "churn label" dan juga "no internet service" dimana tingkat churn menunjukkan persentase pelanggan yang berhenti menggunakan layanan atau produk dalam jangka waktu tertentu. Dalam konteks churn label, "0" mengindikasikan pelanggan yang tetap bertahan, sedangkan "1" menandakan pelanggan yang memilih untuk berhenti berlangganan. Untuk Cluster 0, yang diberi label "Pelanggan Menengah", 74.9% dari pelanggan tetap setia dan memilih untuk tidak berhenti berlangganan. Sebaliknya, Cluster 1, yang diberi label "Pelanggan Pemula", memiliki 65.6% pelanggan yang memutuskan untuk bertahan. Sementara itu, Cluster 2, yang dijuluki "Pelanggan Elite", menunjukkan retensi pelanggan tertinggi dengan 78.4% pelanggan yang memilih untuk tidak berhenti berlangganan.

Dalam menghadapi dinamika pelanggan yang beragam, ada beberapa rekomendasi strategis yang dapat diterapkan oleh perusahaan untuk memastikan retensi pelanggan yang optimal. Pertama, seiring dengan retensi yang menonjol pada kluster "Pelanggan Elite", sangat penting bagi

perusahaan untuk terus fokus pada retensi kluster ini. Mengingat loyalitas mereka yang tinggi, perusahaan sebaiknya memberikan layanan serta penawaran eksklusif bagi kluster ini. Hal ini bisa berupa program loyalitas atau diskon khusus yang dirancang eksklusif untuk memenuhi kebutuhan dan ekspektasi mereka.

Selanjutnya, bagi "Pelanggan Menengah" dan "Pelanggan Pemula", ada potensi besar yang belum sepenuhnya dimanfaatkan. Untuk kluster ini, strategi retensi yang lebih spesifik dan disesuaikan dengan preferensi mereka perlu diterapkan. Ini mungkin mencakup pemberian insentif melalui program referral, atau dengan menawarkan paket bundling produk dengan harga yang lebih terjangkau.

Lebih jauh, untuk mengatasi isu churn yang mungkin terjadi, sangat penting bagi perusahaan untuk mendalami alasan di balik keputusan pelanggan untuk berhenti berlangganan. Metode seperti survei kepuasan pelanggan atau wawancara mendalam dapat dilakukan untuk mengumpulkan insight yang berharga terkait alasan churn.

Terakhir, mengingat betapa pentingnya retensi pelanggan dalam setiap kluster, meningkatkan kualitas layanan pelanggan menjadi prioritas utama. Ini termasuk menyediakan dukungan teknis yang responsif, serta menangani keluhan dan masalah pelanggan dengan cepat dan efisien. Dengan pendekatan ini, perusahaan dapat membangun hubungan yang lebih kuat dengan pelanggan dan memastikan retensi yang berkelanjutan.

Kemudian ada Label "No Internet Service", yang menunjukkan bahwa pelanggan tidak dapat terhubung ke internet karena tidak ada layanan internet tersedia di area mereka. Ini mungkin disebabkan oleh beberapa alasan, seperti kurangnya infrastruktur, biaya yang tinggi, atau area geografis yang sulit dijangkau oleh penyedia layanan internet. Dalam konteks ini, beberapa pelanggan mungkin tidak memiliki akses ke produk atau layanan yang memerlukan koneksi internet. Misalnya, dalam Cluster 0, 22.7% pelanggan tidak memiliki akses internet untuk produk game dan produk musik. Hal ini memberikan gambaran tentang bagian demografi pelanggan yang mungkin menghadapi hambatan dalam mengakses layanan tertentu karena keterbatasan konektivitas.

Label "No Internet Service" mencerminkan sebuah tantangan yang dihadapi oleh sejumlah pelanggan, terutama mereka yang

berada di area dengan keterbatasan infrastruktur internet. Mengingat signifikansi dari keterbatasan ini, sangat penting bagi perusahaan untuk segera mengambil tindakan. Salah satu langkah yang bisa diambil adalah dengan meningkatkan investasi dalam pembangunan dan peningkatan infrastruktur internet di area yang kurang dilayani. Melalui kerjasama strategis dengan pemerintah setempat atau penyedia telekomunikasi lokal, perusahaan dapat memahami dan menangani hambatan geografis dan biaya yang terkait.

Sementara itu, sebagai solusi jangka pendek, perusahaan bisa mempertimbangkan untuk mengembangkan produk atau layanan yang tidak memerlukan koneksi internet, memungkinkan pelanggan di area terbatas untuk tetap mendapatkan nilai dari produk atau layanan yang ditawarkan. Selain itu, meluncurkan program edukasi akan membantu pelanggan memahami manfaat dari koneksi internet, mempersiapkan mereka untuk bertransisi ketika infrastruktur sudah ditingkatkan. Dan untuk menjamin adopsi yang cepat di area yang baru mendapatkan akses, perusahaan bisa menawarkan paket dengan diskon atau subsidi.

Terakhir, mendengarkan suara pelanggan melalui survei akan memberikan wawasan lebih lanjut tentang bagaimana memenuhi kebutuhan mereka dengan lebih efektif. Sehingga, dengan pendekatan holistik ini, perusahaan dapat memastikan bahwa seluruh pelanggan mendapatkan pengalaman yang optimal, terlepas dari keterbatasan geografis mereka.

2) Churn Prediction

Setelah dilakukan proses modeling, didapat hasil sebagai berikut :

Model	Accuracy	AUC-ROC	Macro Precision	Macro Recall	Macro F1-score
Logistic Regression	0.7864	0.7422	0.7376	0.7422	0.7398
Decision Tree	0.7402	0.6941	0.6847	0.6941	0.6888
Random Forest	0.7637	0.7143	0.7102	0.7143	0.7122
XGBoost	0.7921	0.7477	0.7444	0.7477	0.746
LightGBM	0.7942	0.7477	0.7469	0.7477	0.7473
CatBoost	0.7878	0.7356	0.7389	0.7356	0.7372
Gradient Boosting Machine	0.7864	0.7558	0.7395	0.7558	0.7463
AdaBoost	0.7764	0.7511	0.7299	0.7511	0.7379
k-NN	0.7033	0.6903	0.663	0.6903	0.6674
Support Vector Machine	0.7331	0.7405	0.7017	0.7405	0.7062
Naive Bayes	0.7658	0.7459	0.7204	0.7459	0.729
ensemble_model_hard	0.7935	0.7441	0.746	0.7441	0.7451
ensemble_model_soft	0.7928	0.7482	0.7452	0.7482	0.7467

Gambar 3. Hasil Churn Prediction

Dari gambar diatas dapat dilihat, beberapa model machine learning menunjukkan hasil yang cukup baik dalam kinerjanya, sementara beberapa lainnya menunjukkan hasil yang kurang memuaskan.

LightGBM menempati posisi puncak dengan akurasi tertinggi, yaitu 0.7942. Ini menunjukkan

bahwa model ini paling tepat dalam mengidentifikasi kedua kelas data dengan benar. Kemudian, AUC-ROC sebesar 0.7477 menegaskan bahwa model ini memiliki kemampuan yang kuat dalam membedakan antara pelanggan yang akan "churn" dan yang tidak. Terakhir Nilai F1-score yang tinggi menunjukkan bahwa model memiliki keseimbangan yang baik antara precision dan recall.

Dari hasil tersebut penulis merekomendasikan menggunakan model LightGBM untuk melakukan churn prediction karena memiliki performa yang sangat baik. Selain itu juga LightGBM adalah framework pohon berbasis gradien yang dirancang untuk kecepatan dan efisiensi, membuatnya ideal untuk dataset besar.

Namun, meskipun LightGBM telah menunjukkan performa yang luar biasa, selalu ada ruang untuk perbaikan dan optimasi lebih lanjut. Beberapa saran untuk meningkatkan akurasi model di masa depan meliputi:

Feature Engineering: Analisis lebih lanjut pada fitur-fitur yang digunakan mungkin dapat membantu dalam meningkatkan kinerja model. Menciptakan fitur-fitur baru, mengubah fitur yang ada, atau bahkan menghapus fitur yang kurang relevan mungkin dapat meningkatkan hasil prediksi.

Optimasi Hyperparameter: Meskipun LightGBM memiliki performa yang baik, optimasi hyperparameter bisa menjadi cara untuk meningkatkan kinerjanya lebih lanjut. Teknik seperti Grid Search atau Randomized Search dapat digunakan untuk mencari kombinasi hyperparameter terbaik.

Penggunaan Data Tambahan: Memperoleh data tambahan atau memanfaatkan teknik augmentasi data bisa membantu model dalam memahami variasi data dengan lebih baik dan meningkatkan generalisasi.

Penerapan Teknik Regularisasi: Untuk menghindari overfitting, teknik regularisasi seperti L1 atau L2 regularization bisa diaplikasikan.

IV. Kesimpulan

1) Segmentasi Pelanggan:

Analisis segmentasi pelanggan telah berhasil mengidentifikasi tiga kluster utama: "Pelanggan Elite", "Pelanggan Menengah", dan "Pelanggan Pemula". Setiap kluster memiliki karakteristik dan kebutuhan yang unik.

"Pelanggan Elite" adalah kelompok yang memberikan kontribusi besar terhadap pendapatan perusahaan. Meskipun mereka menunjukkan

tingkat loyalitas yang tinggi, faktor churn masih menjadi perhatian utama.

"Pelanggan Menengah" dan "Pelanggan Pemula" memiliki potensi besar untuk kontribusi lebih lanjut. Strategi retensi yang dirancang khusus untuk kluster ini dapat meningkatkan retensi pelanggan.

Pelanggan dengan label "No Internet Service" menghadapi keterbatasan dalam mengakses layanan yang memerlukan koneksi internet, dengan 22.7% dari Cluster 0 tidak memiliki akses ke produk game dan musik. Strategi yang dirancang khusus untuk mengatasi hambatan konektivitas ini dapat membantu dalam memenuhi kebutuhan pelanggan ini.

2) Prediksi Churn:

Beberapa model machine learning telah diuji untuk tujuan prediksi churn. LightGBM menonjol sebagai model dengan performa terbaik, dengan akurasi tertinggi sebesar 0.7942 dan AUC-ROC sebesar 0.7477.

Meskipun LightGBM telah menunjukkan performa yang memuaskan, optimasi lebih lanjut melalui teknik seperti feature engineering, optimasi hyperparameter, dan penggunaan data tambahan dapat meningkatkan kinerjanya.

Disarankan untuk menggunakan LightGBM sebagai model utama dalam melakukan prediksi churn, namun tetap mempertimbangkan potensi model lain atau teknik ensemble untuk hasil yang lebih baik.

Secara umum, pemahaman mendalam tentang segmen pelanggan dan prediksi churn sangat penting untuk menginformasikan strategi retensi pelanggan. Dengan memahami kebutuhan dan perilaku pelanggan yang berbeda, perusahaan dapat merancang intervensi yang tepat untuk meningkatkan kepuasan pelanggan dan meminimalkan kehilangan pelanggan. Dalam konteks bisnis, memaksimalkan retensi pelanggan sambil meminimalkan churn akan membantu dalam mempertahankan pendapatan dan meningkatkan profitabilitas jangka panjang.

DAFTAR PUSTAKA

- [1] Ye, L., Qiu-ru, C., Hai-xu, X., Yi-jun, L., & Zhi-min, Y. (2012). Telecom customer segmentation with K-means clustering. 2012 7th International Conference on Computer Science & Education (ICCSE). doi:10.1109/iccse.2012.6295158
- [2] Qiu, Y., Chen, P., Lin, Z., Yang, Y., Zeng, L., & Fan, Y. (2020). Clustering Analysis for Silent Telecom Customers Based on K-means++. 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). doi:10.1109/itnec48623.2020.9084976
- [3] Jinghua Zhao, Wenbo Zhang, & Yanwei Liu. (2010). Improved K-Means cluster algorithm in telecommunications enterprises customer segmentation. 2010 IEEE International Conference on Information Theory and Information Security. doi:10.1109/icitis.2010.5688749
- [4] Ye, L., Qiu-ru, C., Hai-xu, X., Yi-jun, L., & Zhi-min, Y. (2012). Telecom customer segmentation with K-means clustering. 2012 7th International Conference on Computer Science & Education (ICCSE). doi:10.1109/iccse.2012.6295158
- [5] Bhuse, P., Gandhi, A., Meswani, P., Muni, R., & Katre, N. (2020). Machine Learning Based Telecom-Customer Churn Prediction. 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). doi:10.1109/iciss49785.2020.9315951
- [6] Yabas, U., Cankaya, H. C., & Ince, T. (2012). Customer Churn Prediction for Telecom Services. 2012 IEEE 36th Annual Computer Software and Applications Conference. doi:10.1109/compsac.2012.54
- [7] Wu, Shuli & Yau, Wei-Chuen & Ong, Thian-Song & Chin, Chong. (2021). Integrated Churn Prediction and Customer Segmentation Framework for Telco Business. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3073776.
- [8] Maria, A. (2018, July 20). Role of Machine Learning in Telecoms Industry - Andrea Maria - Medium. Medium. <https://medium.com/@andreausa/role-of-machine-learning-in-telecoms-industry-7bfac18f2fa1>
- [9] Dmello, C. (2022, March 22). Guide On Customer Churn: Don't Just Predict, Prevent it! Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2022/02/guide-on-customer-churn-dont-just-predict-prevent-it/>
- [10] What is market segmentation? Benefits and examples | Kameleoon. (n.d.). Kameleoon. <https://www.kameleoon.com/en/blog/segmentation-audience>
- [11] Rendorio, I. (2023, May 3). What Is Customer Churn Prediction & How to Improve It. Podium. <https://www.podium.com/article/what-is-customer-churn-prediction-how-to-improve-it/>
- [12] How Costly Is Customer Churn in the Telecom Industry? - The European Business Review. (2021, May 20). The European Business Review. <https://www.europeanbusinessreview.com/how-costly-is-customer-churn-in-the-telecom-industry/>
- [13] Scroxton, A. (2018, August 13). How churn is breaking the telecoms market – and what service providers can do about it. The Full Spectrum. <https://www.computerweekly.com/blog/The-Full-Spectrum/How-churn-is-breaking-the-telecoms-market-and-what-service-providers-can-do-about-it>
- [14] What is Customer Churn Prediction? Benefits and Challenges. (2023, August 13). Staircase AI | Customer Intelligence Platform. <https://staircase.ai/learn/churn-prediction/>