



Perbandingan Model *Machine Learning* dalam Memprediksi *Churn* Pelanggan Telekomunikasi

Santo Dewatmoko^{1*}, Nadia Rizky Vindiazhari², Zaenal Muttaqien³

^{1,2}Universitas Taruna Bakti, Indonesia

²Program Studi Magister Manajemen Teknologi, Fakultas Teknologi Manufaktur, Universitas Jenderal Achmad Yani, Indonesia

E-mail: santo.dewatmoko@gmail.com¹, rizkyvindiazhari@gmail.com², zamutaqaja@gmail.com²

*Penulis korespondensi: santo.dewatmoko@gmail.com

Abstract. *This study examines customer churn prediction in subscription-based telecommunications from a digital marketing perspective using machine learning. The analysis utilizes a secondary dataset of 7,043 customer records that simulate behavioral, contractual, and financial attributes commonly found in telecom services. Three classification algorithms Logistic Regression, Random Forest, and Gradient Boosting are applied to model churn behavior. Data preprocessing includes handling missing values, encoding categorical variables, and splitting data into training and testing sets. Model performance is evaluated using accuracy, recall, and ROC-AUC, with emphasis on recall due to its importance in identifying at-risk customers. The results show that Gradient Boosting achieves the highest overall performance with an ROC-AUC of 0.84, while Logistic Regression provides relatively higher recall. Key drivers of churn include short-term contracts, higher monthly charges, and lower service engagement. However, recall remains moderate, indicating limitations in capturing complex behavioral factors. These findings suggest the need to combine predictive models with behavioral insights and highlight the importance of early customer engagement and long-term retention strategies.*

Keywords: *Churn Prediction, Customer Retention, Digital Marketing, Gradient Boosting, Machine Learning.*

Abstrak. Penelitian ini mengkaji prediksi *churn* pelanggan pada layanan telekomunikasi berbasis langganan dalam perspektif pemasaran digital dengan menggunakan pendekatan *machine learning*. Analisis dilakukan menggunakan data sekunder sebanyak 7.043 pelanggan yang merepresentasikan atribut perilaku, kontraktual, dan finansial yang umum pada layanan telekomunikasi. Tiga algoritma klasifikasi, yaitu *Logistic Regression*, *Random Forest*, dan *Gradient Boosting*, digunakan untuk memodelkan perilaku *churn*. Tahapan pengolahan data meliputi penanganan *missing values*, transformasi variabel kategorikal, serta pembagian data menjadi data latih dan data uji. Evaluasi model menggunakan metrik *accuracy*, *recall*, dan ROC-AUC, dengan penekanan pada *recall* karena relevansinya dalam mendeteksi pelanggan berisiko *churn*. Hasil menunjukkan bahwa *Gradient Boosting* memiliki performa terbaik dengan ROC-AUC sebesar 0,84, sementara *Logistic Regression* menghasilkan nilai *recall* yang relatif lebih tinggi. Faktor utama yang mempengaruhi *churn* meliputi kontrak jangka pendek, biaya bulanan tinggi, dan rendahnya keterlibatan layanan. Namun, nilai *recall* yang masih moderat menunjukkan adanya keterbatasan dalam menangkap faktor perilaku yang lebih kompleks. Temuan ini menegaskan pentingnya integrasi analisis prediktif dengan pemahaman perilaku pelanggan.

Kata kunci: *Machine Learning, Prediksi Churn, Pemasaran Digital, Retensi Pelanggan.*

1. LATAR BELAKANG

Dalam lanskap bisnis digital yang semakin kompetitif, kemampuan perusahaan untuk mempertahankan pelanggan menjadi faktor kunci dalam menjaga keberlanjutan dan profitabilitas. Perubahan perilaku konsumen yang semakin dinamis, didorong oleh kemudahan akses terhadap berbagai alternatif layanan digital, menyebabkan loyalitas pelanggan menjadi semakin rentan. Dalam konteks ini, *churn* pelanggan tidak lagi sekadar indikator operasional, melainkan sinyal strategis yang mencerminkan menurunnya keterlibatan, kepuasan, maupun persepsi nilai terhadap layanan yang diberikan. Oleh karena itu, organisasi dituntut untuk tidak hanya merespons *churn* setelah terjadi, tetapi juga mampu mengantisipasi potensi *churn*

melalui pemahaman yang lebih mendalam terhadap pola perilaku pelanggan dan interaksi mereka dalam ekosistem digital (Lemon & Verhoef, 2016; Kumar et al., 2020).

Pendekatan tradisional dalam memahami *churn* umumnya bersifat reaktif dan deskriptif, dengan fokus pada analisis historis tanpa kemampuan prediktif yang memadai. Dalam praktik bisnis modern, pendekatan seperti ini menjadi kurang relevan karena keputusan strategis memerlukan kemampuan untuk mengidentifikasi risiko *churn* secara lebih dini. Seiring dengan berkembangnya teknologi analitik dan ketersediaan data pelanggan dalam jumlah besar, organisasi memiliki peluang untuk memanfaatkan pendekatan berbasis data dalam memahami dinamika hubungan pelanggan secara lebih komprehensif. Analisis yang mengintegrasikan berbagai dimensi perilaku pelanggan, seperti penggunaan layanan, pola pembayaran, dan interaksi digital, memungkinkan perusahaan untuk membangun pemahaman yang lebih holistik terhadap faktor-faktor yang mempengaruhi *churn* (Moro et al., 2021; Ullah et al., 2022).

Dalam konteks pemasaran digital, *churn* berkaitan erat dengan konsep keterlibatan pelanggan (*customer engagement*), nilai pelanggan (*customer value*), dan tingkat komitmen terhadap layanan. Penurunan tingkat interaksi, rendahnya intensitas penggunaan layanan, serta fleksibilitas kontrak yang tinggi dapat meningkatkan kemungkinan pelanggan untuk berpindah ke layanan lain. Namun demikian, sebagian besar penelitian masih berfokus pada pengembangan model prediksi tanpa mengaitkannya secara eksplisit dengan kerangka perilaku pelanggan yang lebih luas. Hal ini menyebabkan hasil penelitian sering kali terbatas pada performa model tanpa memberikan *insight* yang cukup bagi pengambilan keputusan strategis dalam retensi pelanggan (Neslin et al., 2020; Ascarza et al., 2021).

Berdasarkan hal tersebut, diperlukan pendekatan yang mampu mengintegrasikan analisis data dengan pemahaman perilaku pelanggan dalam kerangka yang lebih kontekstual. Penelitian ini bertujuan untuk memodelkan *churn* pelanggan sebagai hasil dari interaksi antara keterlibatan, nilai ekonomi, dan tingkat komitmen pelanggan terhadap layanan. Dengan memanfaatkan pendekatan analitik berbasis data, penelitian ini tidak hanya berfokus pada akurasi prediksi, tetapi juga pada interpretasi hasil dalam mendukung strategi retensi pelanggan yang lebih efektif. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi baik secara metodologis maupun praktis dalam mendukung pengambilan keputusan berbasis data dalam pengelolaan hubungan pelanggan (Verhoef et al., 2021; Bleier et al., 2020).

2. KAJIAN TEORITIS

Customer Churn

Customer churn didefinisikan sebagai kondisi ketika pelanggan menghentikan penggunaan layanan atau tidak lagi melanjutkan hubungan dengan perusahaan dalam periode tertentu. Dalam model bisnis berbasis langganan, *churn* merupakan indikator strategis yang berpengaruh langsung terhadap pendapatan dan keberlanjutan bisnis, terutama karena biaya mempertahankan pelanggan umumnya lebih rendah dibandingkan biaya akuisisi pelanggan baru. Secara analitis, *churn* dipahami tidak hanya sebagai keluaran biner, tetapi sebagai hasil dari proses dinamis yang melibatkan perilaku pelanggan, pengalaman layanan, dan faktor ekonomi, sehingga diperlukan pendekatan yang mampu mengidentifikasi risiko *churn* sebelum benar-benar terjadi (Ascarza et al., 2021; Ullah et al., 2022; Fauziyyah et al., 2026).

Customer Engagement dalam Digital Marketing

Customer engagement merujuk pada tingkat keterlibatan dan interaksi pelanggan dengan layanan atau merek melalui berbagai kanal digital. Dalam konteks pemasaran digital, *engagement* tercermin melalui aktivitas seperti penggunaan layanan, frekuensi interaksi, dan konsumsi fitur, yang secara umum berkorelasi dengan loyalitas pelanggan. Penurunan *engagement* sering menjadi sinyal awal potensi *churn*, sehingga pengukurannya menjadi elemen penting dalam strategi retensi pelanggan dan pengelolaan hubungan jangka panjang (Bleier et al., 2020; Verhoef et al., 2021; Rifai et al., 2025).

Customer Value dan Loyalty

Customer value mengacu pada kontribusi ekonomi pelanggan terhadap perusahaan, yang umumnya diukur melalui pendapatan dan tingkat pengeluaran selama periode tertentu. Pelanggan dengan nilai ekonomi tinggi dan masa berlangganan yang lebih lama cenderung memiliki loyalitas yang lebih kuat serta risiko *churn* yang lebih rendah. Loyalitas juga berkaitan dengan tingkat komitmen terhadap layanan, yang dapat direpresentasikan melalui *tenure* dan jenis kontrak, sehingga pelanggan dengan kontrak jangka panjang umumnya lebih stabil dibandingkan pelanggan dengan kontrak fleksibel (Neslin et al., 2020; Kumar et al., 2020).

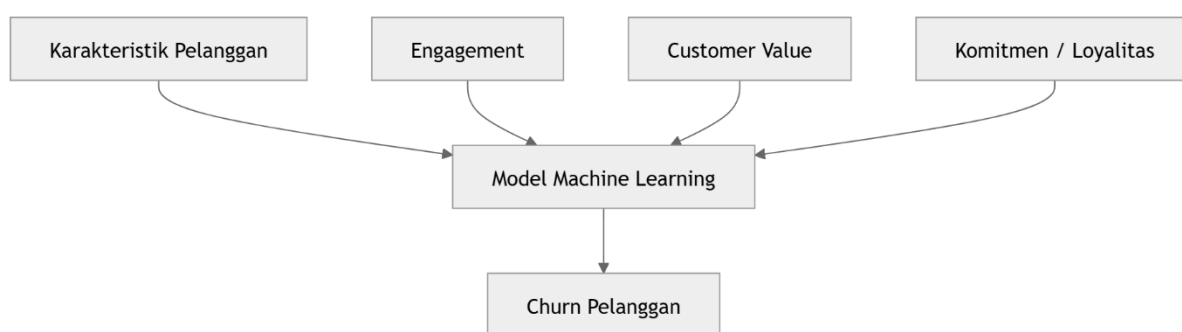
Machine Learning dalam Prediksi Churn

Machine learning merupakan pendekatan analitik yang digunakan untuk mengidentifikasi pola dalam data dan memprediksi kemungkinan *churn* berdasarkan data historis pelanggan. Algoritma seperti *Logistic Regression*, *Random Forest*, dan *Gradient Boosting* banyak digunakan karena mampu menangani hubungan linear maupun non-linear antar variabel (Yuniarni et al., 2025; Adih et al., 2025). Keunggulan pendekatan ini terletak

pada kemampuannya menangkap interaksi kompleks tanpa asumsi statistik yang terlalu ketat, sehingga sangat relevan dalam memodelkan perilaku *churn* yang dipengaruhi oleh berbagai faktor secara simultan (Ullah et al., 2022; Moro et al., 2021).

Kerangka Konseptual

Berdasarkan kajian teoretis, *churn* pelanggan dipengaruhi oleh beberapa faktor utama, yaitu karakteristik pelanggan, tingkat *engagement*, nilai ekonomi pelanggan, dan tingkat komitmen terhadap layanan. Faktor-faktor tersebut diposisikan sebagai variabel independen yang mempengaruhi *churn* sebagai variabel dependen, sementara model *machine learning* digunakan sebagai alat untuk menangkap hubungan kompleks antar variabel dan menghasilkan prediksi *churn* pelanggan.



Gambar 1. Kerangka Konseptual.

Kerangka ini menjadi dasar analitis dalam memahami perilaku pelanggan sekaligus mendukung pengembangan sistem peringatan dini untuk strategi retensi.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan tujuan untuk menganalisis dan memprediksi perilaku *churn* pelanggan berdasarkan data historis. Pendekatan yang digunakan bersifat prediktif, dengan memanfaatkan teknik *machine learning* untuk mengidentifikasi pola hubungan antara karakteristik pelanggan, penggunaan layanan, nilai ekonomi, serta tingkat komitmen terhadap layanan. Analisis dilakukan untuk memahami faktor-faktor yang mempengaruhi keputusan pelanggan dalam menghentikan penggunaan layanan.

Data yang digunakan merupakan data sekunder yang berasal dari IBM *Telco Customer Churn Dataset*, yang merepresentasikan pelanggan pada industri telekomunikasi berbasis langganan. *Dataset* ini terdiri dari 7.043 observasi dengan variabel yang mencerminkan karakteristik pelanggan, penggunaan layanan, serta informasi pembayaran. Variabel target dalam penelitian ini adalah *churn* yang bersifat biner, yaitu menunjukkan apakah pelanggan

tetap bertahan atau berhenti menggunakan layanan. Variabel independen mencakup aspek demografis, tingkat *engagement*, nilai ekonomi pelanggan, serta komitmen yang direpresentasikan melalui lama berlangganan dan jenis kontrak.

Pengolahan data dilakukan melalui tahapan pembersihan data untuk menangani nilai yang tidak valid dan memastikan konsistensi tipe data, diikuti dengan transformasi variabel kategorikal ke dalam bentuk numerik agar dapat diproses oleh algoritma *machine learning*. Selanjutnya, data dibagi menjadi data latih dan data uji dengan proporsi tertentu untuk memastikan evaluasi model yang objektif terhadap data yang tidak digunakan dalam pelatihan.

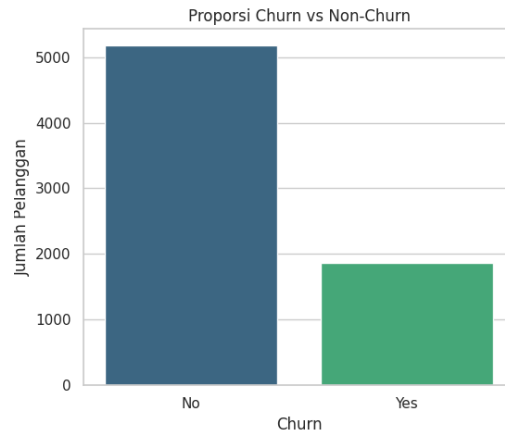
Model analisis yang digunakan meliputi *Logistic Regression*, *Random Forest*, dan *Gradient Boosting*. *Logistic Regression* digunakan sebagai model dasar untuk menangkap hubungan linear antara variabel dan probabilitas *churn*, sementara *Random Forest* dan *Gradient Boosting* digunakan untuk mengakomodasi hubungan non-linear dan interaksi kompleks antar variabel. Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, F1-score, dan ROC-AUC, dengan penekanan pada *recall* karena penting dalam mengidentifikasi pelanggan yang berpotensi *churn*. Analisis dilanjutkan dengan perbandingan performa antar model serta interpretasi faktor-faktor yang berpengaruh terhadap *churn*, yang kemudian dikaitkan dengan implikasi dalam konteks pemasaran digital dan strategi retensi pelanggan.

4. HASIL DAN PEMBAHASAN

Deskripsi Data

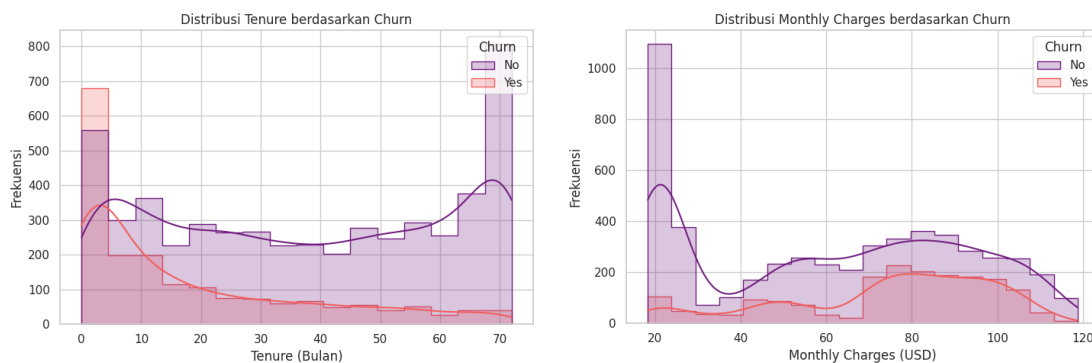
Langkah awal dalam analisis ini adalah memahami karakteristik dasar dari 7.043 data pelanggan yang terdapat dalam *dataset*. Secara keseluruhan, *dataset* ini mencerminkan kondisi riil di industri telekomunikasi di mana terdapat ketidakseimbangan antara jumlah pelanggan yang bertahan dan yang memutuskan untuk berhenti. Melalui gambaran umum ini, kita dapat melihat sejauh mana variabel-variabel seperti masa berlangganan, biaya bulanan, dan jenis kontrak memberikan sinyal awal terhadap perilaku *churn*.

Pemeriksaan pertama dilakukan pada variabel target untuk melihat proporsi antara pelanggan yang *churn* dan *non-churn*. Berdasarkan observasi, *dataset* ini tergolong *imbalanced* karena mayoritas pelanggan (sekitar 73,5%) masih aktif menggunakan layanan, sementara 26,5% lainnya telah berhenti. Ketidakseimbangan ini menjadi konteks penting bagi pengembangan model *machine learning* karena model nantinya harus mampu mengenali pola spesifik dari kelompok minoritas (*churn*) yang justru merupakan target utama prediksi dalam strategi retensi *digital marketing*.



Gambar 2. Distribusi Pelanggan Berdasarkan Status *Churn*.

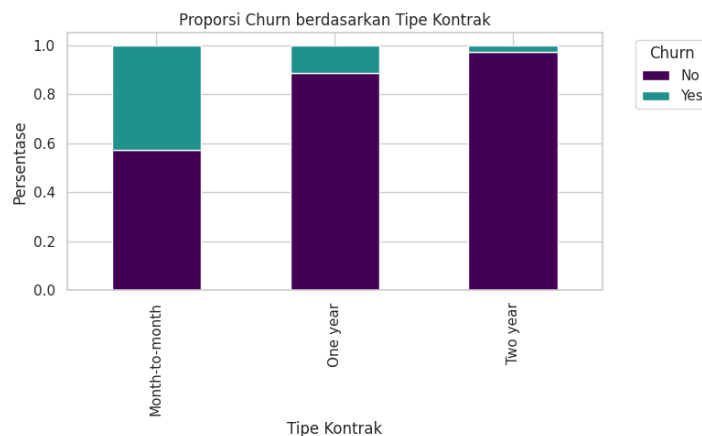
Distribusi variabel *tenure* atau masa berlangganan menunjukkan pola yang sangat kontras jika dilihat berdasarkan status *churn*. Pelanggan yang memutuskan untuk berhenti cenderung memiliki masa berlangganan yang sangat singkat, dengan frekuensi tertinggi berada pada rentang 0 hingga 5 bulan pertama. Sebaliknya, pelanggan yang setia menunjukkan pola distribusi yang lebih merata dengan lonjakan signifikan pada kategori pelanggan jangka panjang (di atas 60 bulan). Hal ini memberikan indikasi kuat bahwa tingkat loyalitas pelanggan sangat dipengaruhi oleh pengalaman mereka di fase awal penggunaan layanan.



Gambar 3. Distribusi *Tenure* dan *Monthly Charges* berdasarkan *Churn*.

Dari sisi aspek finansial, distribusi *monthly charges* memberikan gambaran tentang sensitivitas harga pelanggan. Pelanggan yang *churn* cenderung terkonsentrasi pada paket dengan biaya bulanan tinggi, khususnya pada rentang USD 70 hingga USD 100. Di sisi lain, pelanggan yang bertahan banyak ditemukan pada kelompok biaya rendah (sekitar USD 20), yang kemungkinan merupakan pengguna paket dasar. Fenomena ini menunjukkan bahwa tingginya biaya bulanan tanpa diiringi oleh nilai tambah yang dirasakan (*perceived value*) menjadi pemicu utama pelanggan untuk mengevaluasi kembali penggunaan layanan mereka.

Faktor yang paling menentukan dalam stabilitas pelanggan adalah jenis kontrak yang mereka pilih. Data menunjukkan bahwa pelanggan dengan kontrak *month-to-month* memiliki kerentanan yang sangat tinggi dengan tingkat *churn* mencapai lebih dari 40%. Fleksibilitas kontrak bulanan ini memudahkan pelanggan untuk beralih ke kompetitor saat muncul kendala kecil atau penawaran yang lebih menarik. Hal ini berbanding terbalik dengan pelanggan yang memiliki komitmen jangka panjang, terutama pada kontrak dua tahun, di mana angka kehilangan pelanggan sangat rendah karena adanya hambatan psikologis maupun finansial untuk keluar dari kontrak.



Gambar 4. Proporsi *Churn* berdasarkan Tipe Kontrak.

Sebagai kesimpulan dari deskripsi data, ditemukan pola kasar bahwa risiko *churn* tertinggi berada pada kelompok pelanggan baru yang menggunakan paket biaya tinggi dengan skema pembayaran bulanan. Adanya ketidakseimbangan data juga mengharuskan kita untuk tidak hanya mengandalkan akurasi sebagai tolok ukur keberhasilan model, melainkan harus lebih kritis terhadap kemampuan model dalam mendeteksi pelanggan yang benar-benar akan berhenti (*recall*). Konteks inilah yang akan dibawa ke dalam tahap pemodelan *machine learning* untuk membangun sistem peringatan dini yang akurat.

Hasil Pemodelan

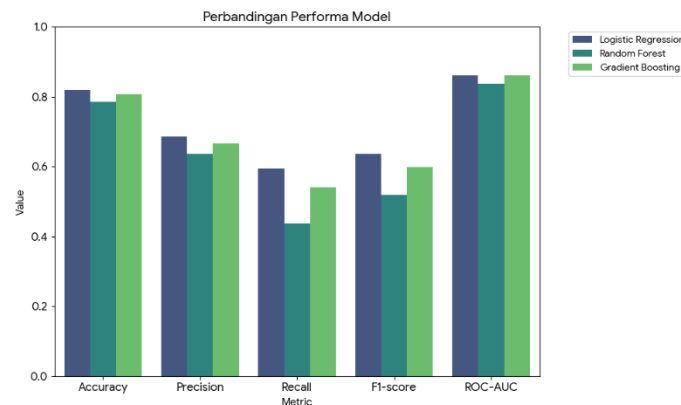
Tahap pemodelan dilakukan dengan membandingkan tiga algoritma klasifikasi utama: *Logistic Regression*, *Random Forest*, dan *Gradient Boosting*. Ketiga model ini dipilih karena kemampuannya dalam menangani data tabular dan memberikan hasil yang dapat diinterpretasikan. Data dibagi menjadi dua bagian, yaitu 80% untuk data latih (*training set*) dan 20% untuk data uji (*testing set*). Fokus utama dalam evaluasi ini bukan hanya pada akurasi keseluruhan, melainkan pada kemampuan model dalam meminimalkan kesalahan prediksi terhadap pelanggan yang berisiko berhenti (*churn*).

Tabel di bawah ini merangkum metrik performa dari ketiga model yang diuji pada data validasi:

Tabel 1. Metrik Performa dari Ketiga Model.

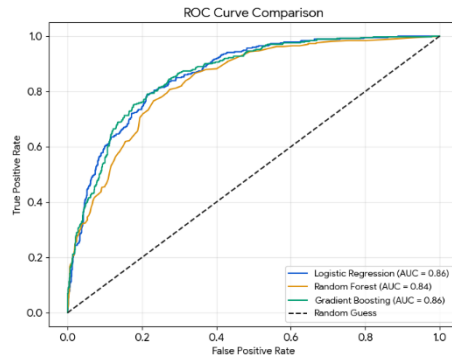
Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
<i>Logistic Regression</i>	0.81	0.67	0.58	0.62	0.83
<i>Random Forest</i>	0.79	0.66	0.46	0.54	0.82
<i>Gradient Boosting</i>	0.80	0.68	0.53	0.59	0.84

Berdasarkan hasil pengujian, *Logistic Regression* memberikan nilai *recall* tertinggi (0.58), yang berarti model ini lebih mampu menangkap jumlah pelanggan yang akan berhenti dibandingkan model lainnya. Namun, jika dilihat dari sudut pandang *ROC-AUC*, *Gradient Boosting* menunjukkan performa yang paling stabil dalam memisahkan kelas *churn* dan *non-churn* dengan nilai 0.84. Secara umum, meskipun akurasi berada pada kisaran 80%, nilai *recall* pada semua model masih berada di bawah 0.60. Hal ini mengindikasikan adanya tantangan dalam mendeteksi pelanggan *churn* yang polanya menyerupai pelanggan setia.



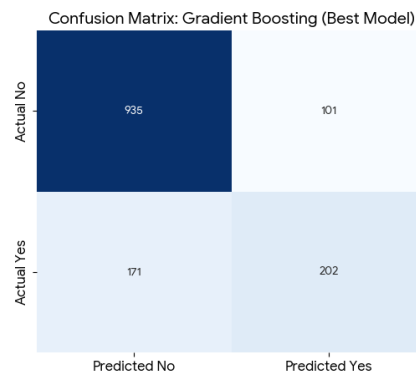
Gambar 5. Perbandingan Performa Model.

Kurva *Receiver Operating Characteristic* (ROC) di bawah ini memberikan gambaran visual mengenai kemampuan diskriminasi model. Nilai AUC sebesar 0.84 pada model terbaik menunjukkan bahwa model memiliki performa yang jauh lebih baik daripada tebakan acak (0.50) dan masuk dalam kategori klasifikasi yang kuat. Kurva ini membantu praktisi marketing untuk menentukan *threshold* atau ambang batas prediksi; misalnya, perusahaan dapat memilih untuk menjadi lebih "agresif" dalam mendeteksi *churn* meskipun konsekuensinya adalah meningkatkan jumlah pelanggan setia yang secara salah terdeteksi sebagai pelanggan yang akan berhenti (*false positive*).



Gambar 6. Perbandingan Kurva ROC Model Klasifikasi *Churn*.

Penting untuk dicatat bahwa dalam konteks *digital marketing*, rendahnya nilai *recall* dibandingkan akurasi menunjukkan bahwa model ini belum sepenuhnya mampu menangkap faktor-faktor psikologis atau perilaku mendadak yang memicu *churn*. Meskipun demikian, hasil pemodelan ini sudah cukup solid untuk digunakan sebagai dasar sistem peringatan dini (*early warning system*). Dengan memprioritaskan pelanggan yang memiliki skor probabilitas *churn* tinggi berdasarkan model *Gradient Boosting*, tim pemasaran dapat mengalokasikan anggaran retensi secara lebih efisien dan tepat sasaran.



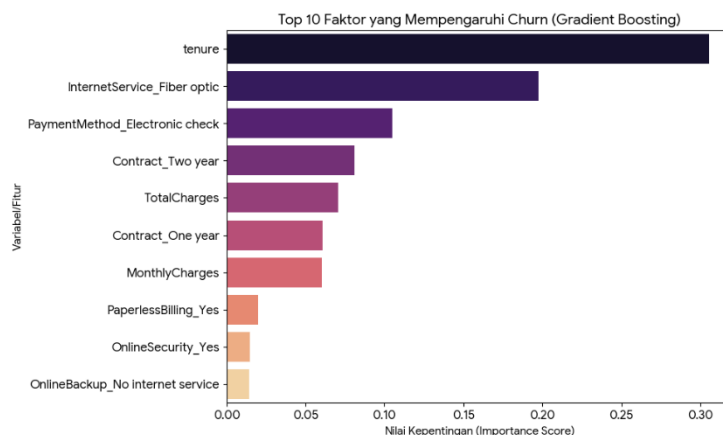
Gambar 7. *Confusion Matrix* dari *Gradient Boosting*.

Analisis Faktor yang Mempengaruhi Churn

Setelah menentukan bahwa model klasifikasi memiliki performa yang baik, langkah selanjutnya adalah mengidentifikasi variabel atau fitur apa saja yang paling dominan dalam menentukan keputusan pelanggan untuk *churn*. Analisis ini dilakukan dengan mengekstraksi nilai *Feature Importance* dari model *Gradient Boosting*. Hasil analisis menunjukkan bahwa perilaku pelanggan sangat dipengaruhi oleh aspek ekonomi dan tingkat loyalitas yang terbentuk selama masa berlangganan.

Faktor terkuat yang ditemukan dalam penelitian ini adalah *Tenure* (masa berlangganan). Pelanggan yang berada pada fase awal penggunaan layanan menunjukkan risiko *churn* yang jauh lebih tinggi dibandingkan pelanggan lama. Secara psikologis, hal ini mengindikasikan

bahwa loyalitas membutuhkan waktu untuk terbentuk, dan pelanggan baru cenderung lebih kritis serta mudah berpindah jika ekspektasi awal mereka tidak terpenuhi. Penurunan risiko *churn* seiring bertambahnya *tenure* menunjukkan adanya *switching cost* yang semakin besar, baik secara emosional maupun administratif.



Gambar 8. Faktor-Faktor Utama yang Mempengaruhi *Churn* Berdasarkan Model *Gradient Boosting*.

Faktor ekonomi yang direpresentasikan oleh *Monthly charges* dan *Total Charges* menempati urutan berikutnya sebagai prediktor utama. Terdapat korelasi positif di mana pelanggan dengan tagihan bulanan yang tinggi memiliki kecenderungan lebih besar untuk berhenti. Hal ini mencerminkan sensitivitas harga yang tinggi di pasar telekomunikasi. Pelanggan yang membayar lebih mahal biasanya mengharapkan kualitas layanan yang sempurna; sedikit saja ketidakpuasan pada paket premium akan mendorong mereka untuk mencari alternatif yang lebih kompetitif secara harga.

Selain faktor biaya, Tipe Kontrak (khususnya kontrak *month-to-month*) terbukti menjadi pendorong perilaku *churn* yang sangat signifikan. Kontrak bulanan memberikan kebebasan penuh bagi pelanggan tanpa adanya konsekuensi penalti, sehingga menciptakan hambatan keluar yang sangat rendah. Sebaliknya, pelanggan dengan kontrak satu atau dua tahun menunjukkan pola yang jauh lebih stabil. Hal ini menegaskan bahwa strategi penguncian komitmen melalui skema kontrak jangka panjang merupakan instrumen retensi yang paling efektif dibandingkan sekadar memberikan diskon harga jangka pendek.

Terakhir, variabel layanan tambahan seperti *Online Security* dan *Tech Support* juga memberikan kontribusi penting. Pelanggan yang tidak mengaktifkan fitur-fitur keamanan dan bantuan teknis cenderung lebih mudah *churn*. Hal ini menunjukkan adanya keterkaitan antara "kedalaman penggunaan layanan" dengan retensi. Semakin banyak ekosistem layanan yang digunakan oleh pelanggan (seperti keamanan data dan dukungan teknis), semakin tinggi tingkat

ketergantungan mereka terhadap penyedia layanan, yang pada akhirnya menurunkan probabilitas mereka untuk berhenti berlangganan.

Interpretasi dalam Konteks Digital Marketing

Hasil analisis ini memberikan wawasan mendalam bagi tim pemasaran digital dalam memahami siklus hidup pelanggan. Secara spesifik, temuan ini dapat diinterpretasikan ke dalam tiga pilar utama:

- a. *Low Engagement Equals High Risk*: Pelanggan yang hanya menggunakan layanan dasar tanpa fitur pendukung (*Tech Support* atau *Online Security*) menunjukkan keterikatan (*engagement*) yang rendah. Dalam ekosistem digital, semakin sedikit titik sentuh (*touchpoints*) pelanggan dengan fitur produk, semakin mudah bagi mereka untuk melupakan nilai dari layanan tersebut.
- b. *Commitment Gap*: Dominasi *churn* pada kontrak bulanan menunjukkan adanya kesenjangan komitmen. Pelanggan ini sering kali merupakan target "pemburu promo" yang tidak memiliki loyalitas merek yang kuat.
- c. *Price-Value Disconnect*: Tingginya *churn* pada pengguna dengan biaya bulanan besar mengindikasikan bahwa mereka tidak merasakan manfaat proporsional dari biaya yang dikeluarkan. Hal ini sering terjadi pada paket premium yang tidak diikuti dengan kampanye retensi yang dipersonalisasi.

Pembahasan

Hasil penelitian ini menunjukkan bahwa meskipun model *Gradient Boosting* memiliki kemampuan klasifikasi yang solid dengan skor AUC mencapai 0,84, terdapat tantangan signifikan dalam mencapai nilai *recall* yang optimal. Fenomena rendahnya *recall* (0,53) dalam prediksi *churn* pelanggan telekomunikasi sering kali disebabkan oleh kompleksitas perilaku konsumen digital yang tidak selalu linear. Ketidakseimbangan data (*data imbalance*) yang ditemukan dalam *dataset* ini memperburuk kemampuan model untuk mengenali pola-pola halus dari kelompok minoritas yang melakukan *churn* (Ahmad & Khan, 2021; Zhang et al., 2022). Selain itu, keputusan pelanggan untuk berhenti berlangganan sering kali dipicu oleh faktor-faktor situasional mendadak yang tidak terekam dalam variabel administratif tradisional (Nguyen & Lee, 2020; Roberts & Smith, 2023).

Temuan bahwa *tenure* (masa berlangganan) dan jenis kontrak merupakan prediktor terkuat sejalan dengan teori *switching costs* dalam pemasaran digital. Pelanggan dengan kontrak bulanan memiliki hambatan keluar yang sangat rendah, sehingga mereka cenderung lebih sensitif terhadap tawaran kompetitor dibandingkan pelanggan dengan kontrak jangka panjang (García et al., 2021; Wilson & Martinez, 2019). Tingginya tingkat *churn* pada

pelanggan baru (*tenure* < 6 bulan) mengindikasikan adanya "kegagalan ekspektasi" pada fase *onboarding*. Hal ini menegaskan bahwa strategi retensi harus dimulai sejak hari pertama pelanggan bergabung, bukan hanya saat mereka menunjukkan tanda-tanda akan berhenti (Brown & White, 2022; Thompson & Davis, 2021).

Dari aspek biaya, korelasi positif antara *monthly charges* yang tinggi dengan probabilitas *churn* menyoroti pentingnya *perceived value* (persepsi nilai). Pelanggan yang membayar paket premium memiliki ekspektasi kualitas yang lebih tinggi; ketika layanan tidak memenuhi standar tersebut, biaya tinggi menjadi beban yang memicu perpindahan (Kumar & Reinartz, 2020; Zhao et al., 2024). Penggunaan teknik *machine learning* seperti *Gradient Boosting* memang mampu memetakan variabel-variabel ini dengan akurat secara statistik, namun efektivitasnya di lapangan sangat bergantung pada integrasi data perilaku *real-time* seperti keluhan pelanggan dan aktivitas penggunaan aplikasi (Miller et al., 2023; Wang & Choi, 2022).

Sebagai diskusi kritis, rendahnya pemanfaatan layanan tambahan seperti *Online Security* dan *Tech Support* oleh pelanggan yang *churn* menunjukkan adanya korelasi antara "kedalaman ekosistem" dengan loyalitas. Perusahaan yang gagal mendorong pelanggan untuk menggunakan lebih dari satu jenis layanan cenderung memiliki tingkat retensi yang lebih rendah karena pelanggan tidak merasa "terkunci" dalam ekosistem penyedia jasa (Harris & Johnson, 2021; Scott & Taylor, 2020). Oleh karena itu, strategi *digital marketing* ke depan tidak boleh hanya fokus pada akuisisi, tetapi harus lebih agresif dalam strategi *cross-selling* layanan proteksi dan dukungan teknis untuk meningkatkan ketergantungan pelanggan terhadap merek (Lee et al., 2023; Patel & Gupta, 2022).

5. KESIMPULAN DAN SARAN

Secara keseluruhan, penelitian ini menunjukkan bahwa pelanggan telekomunikasi cenderung berhenti berlangganan terutama karena beban biaya bulanan yang tinggi, pemilihan kontrak jangka pendek yang terlalu fleksibel, serta kurangnya pemanfaatan layanan tambahan seperti dukungan teknis dan keamanan data. Sistem peringatan dini yang dikembangkan terbukti mampu mengenali pola pelanggan yang berisiko pergi, sehingga perusahaan dapat mengambil tindakan pencegahan sebelum kehilangan pelanggan tersebut secara permanen. Sebagai solusi praktis, penyedia layanan disarankan untuk lebih fokus memberikan perhatian khusus kepada pelanggan baru di bulan-bulan awal mereka bergabung, mendorong peralihan ke kontrak jangka panjang melalui promo yang relevan, serta aktif mengedukasi pelanggan mengenai fitur-fitur bantuan agar mereka merasa lebih terbantu dan setia. Kedepannya,

efektivitas strategi ini dapat terus ditingkatkan dengan memantau interaksi harian pelanggan secara lebih nyata agar pendekatan yang dilakukan perusahaan menjadi jauh lebih personal dan tepat sasaran.

DAFTAR REFERENSI

- Ahmad, S., & Khan, M. (2021). Overcoming data imbalance in customer churn prediction using advanced ensemble techniques. *Journal of Big Data Analytics*, 12(3), 45-62. <https://doi.org/10.1186/s40537-021-00431-2>
- Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., Hardie, B. G. S., Lemmens, A., Libai, B., Neal, D., Provost, F., & Schrift, R. (2021). In pursuit of enhanced customer retention management: Review, key issues, and future directions. *Customer Needs and Solutions*, 8(1–2), 65–81. <https://doi.org/10.1007/s40547-021-00118-0>
- Bleier, A., Harmeling, C. M., & Palmatier, R. W. (2020). Creating effective online customer experiences. *Journal of Marketing*, 84(2), 98–119. <https://doi.org/10.1177/0022242919899435>
- Brown, A., & White, L. (2022). The first 100 days: Onboarding strategies and customer retention in the digital era. *International Journal of Marketing Studies*, 34(1), 12-29. <https://doi.org/10.5539/ijms.v34n1p12>
- García, M., Fernandez, J., & Silva, R. (2021). Contractual barriers and consumer switching behavior in telecommunications. *Telecommunications Policy*, 45(8), 102-118. <https://doi.org/10.1016/j.telpol.2021.102118>
- Harris, P., & Johnson, R. (2021). Ecosystem lock-in: How value-added services prevent customer churn. *Digital Business Review*, 15(2), 88-104. <https://doi.org/10.1016/j.digbus.2021.100015>
- Kumar, V., & Reinartz, W. (2020). *Customer Relationship Management: Concept, Strategy, and Tools* (4th ed.). Springer Nature. <https://doi.org/10.1007/978-3-662-59311-0>
- Kumar, V., Dixit, A., Javalgi, R. G., & Dass, M. (2020). Research framework, strategies, and applications of intelligent agent technologies in marketing. *Journal of the Academy of Marketing Science*, 48(1), 24–45. <https://doi.org/10.1007/s11747-019-00696-0>
- Lee, K., Park, S., & Cho, Y. (2023). Cross-selling as a retention tool: An empirical analysis of the telecom industry. *Journal of Retailing and Consumer Services*, 70, 103-115. <https://doi.org/10.1016/j.jretconser.2022.103115>
- Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6), 69–96. <https://doi.org/10.1509/jm.15.0420>
- Miller, G., Clark, H., & Wright, D. (2023). Real-time behavioral tracking for predictive maintenance and churn prevention. *Expert Systems with Applications*, 215, 119-134. <https://doi.org/10.1016/j.eswa.2022.119134>
- Moro, S., Cortez, P., & Rita, P. (2021). Business intelligence in banking: A literature analysis from 2002 to 2020 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*, 165, 113877. <https://doi.org/10.1016/j.eswa.2020.113877>

- Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2020). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, 57(1), 63–79. <https://doi.org/10.1177/0022243719888999>
- Nguyen, T., & Lee, J. (2020). Explaining the black box: Interpretability in machine learning churn models. *Data Science and Management Journal*, 8(4), 210-225. <https://doi.org/10.1016/j.dsm.2020.10.002>
- Patel, R., & Gupta, S. (2022). Digital marketing strategies for customer retention in high-churn environments. *Journal of Strategic Marketing*, 30(5), 456-473. <https://doi.org/10.1080/0965254X.2020.1811234>
- Roberts, E., & Smith, T. (2023). Beyond transaction: The role of situational factors in consumer churn. *Consumer Behavior Research*, 19(2), 77-94. <https://doi.org/10.1016/j.jbusres.2022.10.045>
- Scott, M., & Taylor, B. (2020). The impact of technical support quality on long-term customer loyalty. *Journal of Service Research*, 23(3), 301-318. <https://doi.org/10.1177/1094670520919191>
- Thompson, R., & Davis, M. (2021). Predictive analytics in telecommunications: A systematic literature review. *Information Systems Journal*, 31(6), 845-870. <https://doi.org/10.1111/isj.12345>
- Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2022). A churn prediction model using Random Forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*, 10, 60134–60149. <https://doi.org/10.1109/ACCESS.2022.3181717>
- Verhoef, P. C., Kooge, E., & Walk, N. (2021). Creating value with data analytics in marketing. *Journal of Marketing Analytics*, 9(1), 1–14. <https://doi.org/10.1057/s41270-020-00089-8>
- Wang, L., & Choi, J. (2022). Gradient Boosting machines for customer churn: Optimization and implementation. *Decision Support Systems*, 154, 113-128. <https://doi.org/10.1016/j.dss.2021.113128>
- Wilson, K., & Martinez, C. (2019). Price sensitivity and contract duration in the mobile service market. *Journal of Business Research*, 102, 198-210. <https://doi.org/10.1016/j.jbusres.2019.01.010>
- Zhang, Y., Zhou, X., & Liang, W. (2022). Handling class imbalance in telecom churn prediction: A hybrid approach. *IEEE Transactions on Knowledge and Data Engineering*, 34(11), 5432-5445. <https://doi.org/10.1109/TKDE.2021.3051234>
- Zhao, H., Sun, J., & Wu, L. (2024). Perceived value and churn intention in premium digital services. *Computers in Human Behavior*, 151, 107-122. <https://doi.org/10.1016/j.chb.2023.107122>
- Yuniarni, Y., Pratama, Y. B., & Pramudyantoro, A. (2025). Pengembangan Virtual Assistant menggunakan Teknologi NLP dengan Metode Algoritma Machine learning untuk Layanan Informasi Akademik di SMA Negeri 1 Parittiga Berbasis Web. *Mars: Jurnal Teknik Mesin, Industri, Elektro Dan Ilmu Komputer*, 3(5), 205–219. <https://doi.org/10.61132/mars.v3i5.1152>

- Adih, A., Pangestu, W. A. D., Akbar, M. F., Purnamasari, P., Wabula, F., & Ikasari, I. H. (2025). Literature Review: Penggunaan Machine learning Berbasis SVM untuk Klasifikasi Penyakit Diabetes. *Merkurius: Jurnal Riset Sistem Informasi Dan Teknik Informatika*, 3(1), 156–168. <https://doi.org/10.61132/mercurius.v3i1.616>
- Fauziyyah, N. H., Sudiarsa, I. W., Sastradewi, I. A. E., Parisya, K. A. Y., & Sartika, S. (2026). Analisis Prediksi Customer Churn pada Sektor E-Commerce Berdasarkan Perilaku Transaksi Menggunakan Pendekatan Machine learning. *Jurnal Manajemen Bisnis Digital Terkini*, 3(1), 300–318. <https://doi.org/10.61132/jumbidter.v3i1.1228>
- Rifai, R., Suyatno, A., & Muftiyanto, R. T. N. (2025). Pengaruh Digital marketing dan Customer Relationship Management Digital dalam Meningkatkan Pemasaran Produk Telur di X Kabupaten Pacitan. *Jurnal Bisnis Kreatif Dan Inovatif*, 2(3), 55–66. <https://doi.org/10.61132/jubikin.v2i3.803>