Jurnal Riset Mahasiswa Matematika
Volume 4 (5) (2025), Pages 142-158
p-ISSN: 2808-1552; e-ISSN: 2808-4926

Analisis Churn Pelanggan Telekomunikasi menggunakan Regresi Logistik Biner

Zahwa Rifsya Pangesti and Corry Sormin *

Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Jambi, Jambi

Abstrak

Peningkatan persaingan dalam industri telekomunikasi mendorong perusahaan untuk mengelola risiko pelanggan berhenti berlangganan (churn) secara lebih efektif. Penelitian ini bertujuan mengidentifikasi faktor-faktor yang memengaruhi kemungkinan pelanggan melakukan churn, dengan fokus pada layanan telepon, layanan internet fiber, dan jenis kontrak. Analisis dilakukan menggunakan regresi logistik biner terhadap 288 pelanggan yang dipilih melalui uji Cochran dan teknik stratified random sampling. Hasil penelitian menunjukkan bahwa layanan internet fiber dan jenis kontrak berpengaruh signifikan terhadap risiko churn. Pelanggan pengguna internet fiber memiliki risiko churn sebesar 3,10 kali lebih tinggi dibandingkan pelanggan non-fiber, sedangkan pelanggan dengan kontrak bulanan (month-to-month) memiliki risiko churn sebesar 7,05 kali dibandingkan pelanggan dengan kontrak jangka panjang. Layanan telepon tidak berpengaruh signifikan terhadap churn. Evaluasi model menunjukkan akurasi sebesar 75,35%, dengan sensitivitas untuk kelas churn sebesar 61,8% dan spesifisitas untuk kelas non-churn sebesar 80,2%. Temuan ini menegaskan bahwa jenis kontrak merupakan faktor dominan dalam menentukan risiko churn pelanggan. Hasil penelitian dapat menjadi dasar bagi perusahaan telekomunikasi dalam merancang strategi retensi yang lebih terarah, khususnya bagi pelanggan kontrak bulanan dan pengguna layanan internet fiber.

Kata Kunci: Regresi Logistik Biner; Churn Pelanggan; Layanan Telekomunikasi

Abstract

Increasing competition in the telecommunications industry requires companies to manage customer churn risk more effectively in order to retain customers. This study aims to identify factors influencing customer churn, focusing on phone service, fiber internet service, and contract type. Binary logistic regression was applied to a sample of 288 customers selected using Cochran's formula and stratified random sampling. The results indicate that fiber internet service and contract type significantly affect churn probability. Customers using fiber internet services have a 3.10 times higher risk of churn compared to non-fiber users, while customers with month-to-month contracts face a 7.05 times higher churn risk than those with long-term contracts. Phone service does not have a significant effect on churn. The model achieves an accuracy of 75.35%, with a churn-class sensitivity of 61.8% and a non-churn specificity of 80.2%. These findings confirm that contract type is the most dominant factor influencing customer churn. The results provide practical insights for telecommunications companies to prioritize retention strategies for month-to-month customers and improve service quality for fiber internet users to reduce churn rates.

Keywords: Binary Logistic Regression; Customer Churn; Telecommunication Services

Copyright © 2025 by Authors, Published by JRMM Group. This is an open access article under the CC BY-SA License (https://creativecommons.org/licenses/by-sa/4.0)

*Corresponding author. E-mail: pangestizahwa18@gmail.com

Submitted: 14 November 2025 Reviewed: 21 Desember 2025 Accepted: 30 Desember 2025
DOI: https://dx.doi.org/10.18860/jrmm.v5i2.38982

1 Pendahuluan

Industri telekomunikasi kini menghadapi persaingan yang sangat ketat karena berkembangnya teknologi, semakin luasnya akses internet, dan banyaknya perusahaan penyedia layanan. Dalam situasi seperti ini, menjaga loyalitas pelanggan menjadi sangat krusial, mengingat biaya akuisisi pelanggan baru umumnya lebih tinggi dibandingkan dengan biaya yang diperlukan untuk mempertahankan pelanggan yang sudah ada [1]. Karena itu, masalah churn pelanggan yaitu ketika pelanggan pindah dari satu penyedia ke penyedia lain bukan hanya soal operasional biasa, melainkan juga persoalan strategis yang bisa berdampak pada pendapatan, kinerja bisnis, dan keuntungan perusahaan [2].

Penelitian-penelitian sebelumnya telah membahas masalah churn di industri telekomunikasi dengan metode statistik dan machine learning. Penelitian terbaru oleh Adiansya [3], membandingkan hasil dari model regresi logistik dan Gradient Boosting dalam memprediksi churn dengan menggunakan data pelanggan. Hasilnya menunjukkan bahwa regresi logistik masih relevan karena kemampuannya dalam memberikan penjelasan yang jelas, meskipun akurasi sedikit lebih rendah dibandingkan metode boosting [3]. Penelitian lain menggunakan regresi logistik, Random Forest, dan SVM untuk data pelanggan telekomunikasi, yang menunjukkan bahwa regresi logistik memiliki kemampuan klasifikasi yang cukup baik dengan akurasi sekitar 79% [4]. Penelitian tersebut menunjukkan bahwa metode regresi logistik masih bisa dipakai karena hasil klasifikasinya cukup bagus dan memiliki kelebihan dalam hal kemudahan dalam memahami hasilnya dibandingkan dengan model-model yang sulit diinterpretasikan. Selain itu, studi mengenai algoritma C4.5 juga membuktikan bahwa mengoptimalkan fitur dan parameter bisa meningkatkan tingkat akurasi secara signifikan, meskipun penelitian ini cenderung hanya fokus pada satu jenis algoritma pembelajaran mesin. [5]. Namun, sebagian besar penelitian tentang *churn* masih fokus pada aspek prediksi, sementara pembahasan mengenai sejauh mana pengaruh setiap variabel layanan memengaruhi risiko *churn* biasanya tidak dijelaskan secara rinci.

Selain itu, banyak penelitian tentang *churn* menunjukkan bahwa data *churn* di industri telekomunikasi biasanya memiliki ketidakseimbangan antar kelas, yaitu jumlah pelanggan yang tidak *churn* jauh lebih banyak dibandingkan pelanggan yang *churn*. Hal ini menyebabkan model pengklasifikasi cenderung memihak kelas yang lebih banyak, sehingga kurang akurat dalam memprediksi *churn* [6]. Di sinilah celah dalam penelitian yang ingin diisi oleh penelitian ini, yaitu kebutuhan akan model yang tidak hanya mampu mengklasifikasikan dengan tepat, tetapi juga memberikan perkiraan risiko yang konsisten dan stabil meskipun data yang digunakan tidak seimbang.

Kebaruan penelitian ini terletak pada peningkatan aspek kontribusi ilmiah dengan memperkuat kebaruan dalam pemilihan variabel dan pendekatan metode penelitian. Berbeda dengan penelitian regresi logistik pada umumnya, penelitian ini secara jelas menggabungkan variabel seperti layanan telepon, internet fiber, dan durasi kontrak untuk menentukan tingkat risiko dengan menggunakan nilai odds ratio. Selain itu, peningkatan kebaruan juga terletak pada penggunaan strategi pemilihan sampel yang lebih ketat, yaitu kombinasi uji Cochran untuk menentukan ukuran sampel yang tepat dan *stratified random sampling* untuk mengatasi masalah ketidakseimbangan data, yang sering kali tidak diperhatikan dalam penelitian sebelumnya. Dengan demikian, kontribusi penelitian ini bukan hanya penambahan kecil saja, melainkan memberikan kerangka kerja yang lebih akurat bagi manajemen dalam mengidentifikasi faktor-faktor utama yang menyebabkan kehilangan pelanggan.

Secara keseluruhan, artikel ini dibagi menjadi beberapa bagian. Bagian 2 menjelaskan metode yang digunakan dalam penelitian, termasuk sumber data, pemetaan variabel biner dan prosedur regresi logistik. Bagian 3 berisi hasil analisis dan pembahasan. Bagian 4 menyajikan kesimpulan hasil penelitian dan dampaknya terhadap strategi mempertahankan pelanggan, serta memberikan saran untuk penelitian lebih lanjut.

2 Metode

Bagian ini menguraikan tahapan analisis yang digunakan untuk mengkaji faktor-faktor yang berpengaruh terhadap peluang churn pelanggan telekomunikasi. Proses dimulai dengan memeriksa dan membersihkan data agar semua variabel yang digunakan memiliki kualitas yang baik. Selanjutnya, variabel respons yaitu status churn pelanggan, dikonversi menjadi bentuk biner sesuai kebutuhan model regresi logistik. Setelah itu dilakukan analisis data awal dan pengujian untuk mengetahui apakah ada hubungan yang terlalu kuat antar variabel prediktor (asumsi multikolinearitas). Setelah memastikan asumsi dasar terpenuhi, dibangun model regresi logistik biner untuk mengukur pengaruh variabel layanan telepon *(phone service)*, layanan internet fiber *(internet service)* dan kontrak terhadap kemungkinan churn pelanggan. Pada tahap akhir, model diuji tingkat signifikansinya, diinterpretasikan nilai odds rationya serta diukur akurasi prediksinya, sehingga diperoleh pemahaman yang jelas mengenai faktor-faktor penyebab *churn* dan bisa digunakan sebagai dasar untuk merekomendasikan strategi retensi pelanggan.

2.1 Sumber Data

Data penelitian berasal dari dataset internal pelanggan telekomunikasi yang terdiri dari 7044 entri. Dari jumlah tersebut, terdapat 1869 pelanggan yang mengalami *churn* dan 5174 pelanggan yang tidak mengalami *churn*, menunjukkan adanya ketidakseimbangan kelas yang cukup besar antara kedua kelompok. Ketidakseimbangan ini berpotensi mengganggu ketepatan analisis. Karena ada ketidakseimbangan antara kelompok *churn* dan *non-churn*, maka dilakukan proses sampling proporsional agar kedua kelompok bisa mewakili secara seimbang dalam analisis. Langkah ini sangat penting agar tidak terjadi bias pada model, karena jika kelompok mayoritas terlalu dominan, hasil analisis bisa tidak akurat serta agar kedua kelompok terwakili secara seimbang dalam analisis.

Dalam penelitian ini, meskipun ada 7.044 observasi, peneliti memilih menggunakan 288 sampel yang direpresentasikan melalui uji Cochran. Menggunakan seluruh data dalam model statistik seperti regresi logistik biner pada populasi besar sering kali membuat p-value terlalu sensitif, sehingga semua variabel terlihat signifikan secara tidak benar, yang disebut Large Sample Bias. Dengan mengambil sampel yang terkontrol melalui uji Cochran pada tingkat kepercayaan 95%, penelitian ini bertujuan untuk mendapatkan estimasi parameter yang lebih stabil dan memiliki kemampuan generalisasi yang lebih baik tanpa mengalami *overfitting*.

2.2 Variabel Penelitian

Penelitian ini melibakan variabel churn, layanan telepon, layanan internet fiber dan tipe kontrak. Karena variabel yang digunakan bersifat kategorikal, data tersebut dikonversi ke dalam bentuk numerik melalui proses pengkodean biner yaitu 1 dan 0 sesuai makna masing-masing variabel. Variabel yang digunakan dalam penelitian ini disajikan sebagai berikut:

Tabel 1: Variabel penelitian

Variabel	Keterangan	Pengkodean	Skala Data
Y	Churn	No = 0; Yes = 1	Nominal
X_1	Layanan Telepon	No = 0; Yes = 1	Nominal
X_2	Layanan Internet	Lainnya = 0; Fiber optic = 1	Nominal
X_3	Kontrak	Lainnya = 0; *month-to-month* = 1	Nominal

2.3 Uji Cochran

Uji Cochran digunakan sebagai dasar untuk menentukan ukuran sampel minimum yang diperlukan sehingga estimasi proporsi dalam suatu populasi memiliki tingkat kepercayaan dan batas

kesalahan (margin of error) yang dapat diterima [7]. Rumus dasar Cochran untuk populasi besar dinyatakan sebagai:

$$n_0 = \frac{Z^2\, p(1-p)}{e^2} \tag{1}$$

dengan:

n_0 : ukuran sampel awal (untuk populasi besar),

Z : nilai statistik Z yang sesuai dengan tingkat kepercayaan,

p : estimasi proporsi yang diharapkan (umumnya digunakan $p = 0{,}5$ jika proporsi sebenarnya tidak diketahui),

e : margin of error yang diizinkan.

Untuk populasi terbatas digunakan koreksi populasi terbatas (finite population correction):

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}} \tag{2}$$

dimana N adalah ukuran populasi sebenarnya.

2.4 *Stratified Random Sampling*

Metode *stratified random sampling* digunakan untuk memastikan bahwa setiap subkelompok dalam populasi memperoleh peluang representasi yang dalam sampel penelitian [8]. Teknik ini dilakukan dengan membagi populasi menjadi beberapa kelompok kecil yang memiliki karakteristik serupa. Setelah kelompok-kelompok tersebut terbentuk, sampel diambil secara acak dari masing-masing kelompok, sehingga setiap anggota populasi memiliki peluang yang sama untuk terpilih.

Peneliti menyadari bahwa reduksi data dari 7.044 menjadi 288 observasi bisa menyebabkan hilangnya informasi kecil yang beragam. Namun, risiko ini diatasi dengan menggunakan teknik *stratified random sampling*. Di tengah ketidakseimbangan kelas, di mana jumlah pelanggan yang tidak *churn* jauh lebih banyak, teknik stratifikasi memastikan bahwa kelompok minoritas, yaitu pelanggan yang *churn*, tetap ada dalam sampel secara proporsional. Hal ini dilakukan agar model tidak hanya mengenali pola dari kelas mayoritas, sehingga kemampuan model dalam mendeteksi *churn* tetap baik meskipun jumlah data berkurang.

Penentuan jumlah sampel dari setiap strata dilakukan dengan memperhatikan proporsi jumlah anggota populasi di setiap strata, yang dihitung menggunakan rumus berikut:

$$n_h = \frac{N_h}{N} \times n \tag{3}$$

di mana n_h adalah jumlah sampel pada strata ke-h, N_h merupakan jumlah elemen dalam strata ke-h, N adalah total populasi, dan n menunjukkan jumlah keseluruhan sampel yang diambil.

2.5 Regresi Logistik Biner

Regresi logistik biner diterapkan ketika variabel dependen bersifat dikotomi, yaitu hanya terdiri atas dua kategori kemungkinan, seperti “ya” dan “tidak” atau “berhasil” dan “gagal”. Model ini digunakan untuk mengestimasi probabilitas terjadinya suatu peristiwa (sukses) berdasarkan nilai variabel bebas yang memengaruhinya. Umumnya, kategori sukses dilambangkan dengan Y = 1, sedangkan kategori gagal dilambangkan dengan Y = 0. Model ini bertujuan untuk memperkirakan peluang terjadinya suatu peristiwa (sukses) berdasarkan nilai variabel bebas yang memengaruhinya [9]. Model regresi logistik biner yang digunakan adalah:

$$\pi(x) = \frac{\exp\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k\right)} \tag{4}$$

dimana k merupakan banyaknya variabel bebas. Untuk memudahkan proses penaksiran parameter dalam model regresi, fungsi peluang $\pi(x)$ pada persamaan tersebut diubah menjadi bentuk logit. Transformasi ini menghasilkan bentuk model regresi logistik sebagai berikut [10]:

$$\pi(x)\left(1+\exp\left(\beta_0+\beta_1x_1+\cdots+\beta_kx_k\right)\right)=\exp\left(\beta_0+\beta_1x_1+\cdots+\beta_kx_k\right) \tag{5}$$

$$\pi(x)+\pi(x)[\exp\left(\beta_0+\beta_1x_1+\cdots+\beta_kx_k\right])=\exp\left(\beta_0+\beta_1x_1+\cdots+\beta_kx_k\right) \tag{6}$$

$$\pi(x)=\exp\left(\beta_0+\beta_1x_1+\cdots+\beta_kx_k\right)\left[1-\pi(x)\right] \tag{7}$$

$$\frac{\pi(x)}{1-\pi(x)}=\exp\left(\beta_0+\beta_1x_1+\cdots+\beta_kx_k\right) \tag{8}$$

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right)=\ln\left[\exp\left(\beta_0+\beta_1x_1+\cdots+\beta_kx_k\right)\right]. \tag{9}$$

$$\text{logit}\left[\pi(x)\right]=\ln\left(\frac{\pi(x)}{1-\pi(x)}\right)=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k \tag{10}$$

persamaan (10) adalah bentuk dari model logit yang merupakan fungsi linear dari masing – masing parameter [11].

Analisis ini dilakukan dengan asumsi dasar regresi logistik biner, termasuk asumsi linearitas logit, di mana terdapat hubungan linear antara variabel independen kontinu (jika ada) dengan logit dari variabel dependen. Terkait dengan *interaction effect* (efek interaksi antar variabel independen), penelitian ini berfokus pada pengaruh utama (*main effects*) dari layanan telepon, internet fiber, dan jenis kontrak. Efek interaksi antar variabel tersebut tidak dianalisis secara mendalam dalam studi ini dan diakui sebagai keterbatasan penelitian. Fokus pada main effects dipilih agar interpretasi koefisien melalui odds ratio tetap sederhana dan langsung dapat diimplementasikan secara praktis oleh manajemen perusahaan.

2.6 Asumsi Multikolinearitas

Asumsi multikolinearitas bertujuan untuk mengetahui apakah ada hubungan linear yang kuat antar variabel independen dalam suatu model [12]. Untuk mengetahui adanya multikolinearitas pada setiap variabel prediktor X_j, digunakan ukuran *Variance Inflation Factor (VIF)* yang dihitung dengan rumus:

$$\text{VIF}_j=\frac{1}{1-R_j^2} \tag{11}$$

dengan R_j^2 merupakan koefisien determinasi hasil regresi variabel X_j terhadap seluruh variabel bebas lainnya. Secara umum, apabila VIF$<$ 10, model dianggap tidak mengalami masalah multikolinearitas yang berarti [13].

2.7 Pengujian Parameter

Pengujian parameter dilakukan untuk mengetahui keterikatan antara variabel prediktor dengan variabel respon. Pengujian ini dapat dilakukan secara simultan terhadap keseluruhan model maupun secara parsial dengan menilai pengaruh masing-masing variabel prediktor secara terpisah.

2.7.1 Uji Simultan

Uji serentak menggunakan uji G digunakan untuk menguji apakah semua variabel prediktor secara simultan memiliki pengaruh yang signifikan terhadap variabel respon pada suatu tingkat signifikansi tertentu dalam model regresi secara keseluruhan [14]. Secara formal, hipotesis yang diuji dapat dituliskan sebagai berikut:

$$H_0:\beta_1=\beta_2=\beta_3=0$$

$$H_1:\text{setidaknya satu }\beta_i\neq0\quad(i=1,2,3)$$

Pada pengambilan keputusan, tolak H_0 jika $G > \chi^2_{\alpha,p}$, artinya jika nilai G yang diperoleh lebih besar dari nilai chi-kuadrat table pada taraf signifikansi maka model dengan variabel bebas dianggap memberikan pengaruh yang signifikan terhadap variabel respon.

2.7.2 Uji Parsial

Pengujian parsial untuk menilai pengaruh masing-masing parameter β_i secara individu. Pengujian ini bertujuan untuk mengetahui apakah suatu variabel prediktor layak dimasukkan ke dalam model atau tidak. Hasil uji parsial menunjukkan apakah penambahan variabel tersebut memberikan pengaruh yang signifikan terhadap model.

Adapun hipotesis yang digunakan untuk setiap variabel adalah sebagai berikut:

$H_0 : \beta_i = 0$, artinya tidak terdapat pengaruh antara variabel prediktor dan variabel respon.

$H_1 : \beta_i \neq 0$, artinya terdapat pengaruh antara variabel prediktor dan variabel respon.

Statistik uji yang digunakan adalah uji Wald yang dirumuskan sebagai:

$$W = \frac{\hat{\beta}_i}{\text{SE}(\hat{\beta}_i)} \tag{12}$$

dengan:

$SE(\hat{\beta}_i)$ standar error dari estimasi koefisien β_i.

$\hat{\beta}_i$ = nilai taksiran dari parameter β_i

Kriteria keputusan yaitu jika nilai $|W_{\text{hit}}| > Z_{(\alpha/2)}$ atau p-value $< \alpha$, maka hipotesis nol ditolak, yang berarti variabel prediktor tersebut berpengaruh signifikan terhadap variabel respon [15].

2.8 Uji Kecocokan Model

Uji kecocokan model dilakukan untuk menilai sejauh mana model mampu menggambarkan data dengan baik, yaitu ketika nilai hasil observasi mendekati atau sesuai dengan nilai yang diprediksi oleh model [16]. Dalam analisis regresi logistik, kelayakan model umumnya diuji menggunakan statistik chi-kuadrat melalui uji Hosmer dan Lemeshow sebagai ukuran tingkat kecocokan model terhadap data [17].

Hipotesis yang digunakan untuk melakukan uji kesesuaian model yaitu:

H_0: model yang dihipotesiskan fit dengan data

H_1: model yang dihipotesiskan tidak fit dengan data

Pengujian kesesuaian model dilakukan dengan menggunakan statistik uji $\hat{C}$ yang mengikuti distribusi chi-kuadrat pada taraf signifikansi α dengan derajat bebas sebesar $g - 2$. Keputusan pengujian ditentukan berdasarkan nilai p-value dan nilai kritis chi-kuadrat. Jika nilai $\hat{C}$ lebih besar dari $\chi^2_{(g-2,\alpha)}$ atau p-value lebih kecil dari α, maka H_0 ditolak, yang berarti model yang terbentuk sudah fit dengan data [18].

2.9 Odds Ratio

Odds ratio menunjukkan seberapa besar kemungkinan suatu kejadian terjadi dan sering digunakan dalam model regresi logistik. Nilai odds ratio dihitung melalui persamaan:

$$\text{OR} = e^{\beta} \tag{13}$$

dimana, β = koefisien regresi dari variabel prediktor.

Jika nilai odds ratio > 1, artinya variabel prediktor meningkatkan kemungkinan terjadinya kejadian pada variabel respon, sedangkan jika nilai odds ratio < 1, variabel prediktor justru mengurangi kemungkinan terjadinya kejadian tersebut. Odds ratio menunjukkan bahwa resiko kemungkinan terjadi y=1 pada kategori x=1 adalah sebesar e^{β_i} kali lebih besar dibandingkan resiko terjadinya y=1 pada kategori x=0 [17].

3 Hasil dan Pembahasan

Bagian ini membahas hasil dari analisis regresi logistik biner yang digunakan untuk mengetahui faktor-faktor apa saja yang berpengaruh terhadap kemungkinan pelanggan telekomunikasi berhenti menggunakan layanan mereka (*churn*). Pembahasan disusun dengan urutan yang jelas, mulai dari penjelasan data secara umum, pengujian kecocokan model, hingga penjelasan koefisien regresi pada variabel-variabel yang diteliti. Setiap tahap dijelaskan secara singkat untuk menunjukkan bagaimana model tersebut dibuat, diuji, dan diartikan dalam menilai dampak dari variabel layanan telepon (*phone service*), layanan internet fiber dan jenis kontrak terhadap peluang *churn* pelanggan telekomunikasi.

3.1 Uji Cochran

Penentuan ukuran sampel menggunakan (1) untuk menghitung jumlah sampel minimum dengan tingkat kepercayaan 95% dan batas kesalahan 5%. Untuk nilai Z=1,96, p=0,2653 dan e=0,05, sehingga diperoleh:

$$n_0 = \frac{(1.96)^2 \times 0.2653\,(1 - 0.2653)}{(0.05)^2} \approx 299.566 \tag{14}$$

Karena populasi penelitian berjumlah 7043 pelanggan, maka ukuran sampel disesuaikan menggunakan (2) sebagai berikut:

$$n = \frac{299.566}{1 + \left(\frac{298.566}{7043}\right)} = 287.383 \approx 288 \tag{15}$$

dengan demikian, diperoleh ukuran sampel minimum sebanyak 288 pelanggan.

3.2 *Stratified Random Sampling*

Setelah ukuran sampel diketahui, pengambilan data dilakukan menggunakan teknik *stratified random sampling*. Dalam penelitian ini, strata yang digunakan adalah *churn* pelanggan (*yes* dan *no*). Diperoleh hasilnya dengan menggunakan (3) yaitu:

$$n_{\text{Yes}} = \frac{1869}{7043} \times 288 \approx 76 \tag{16}$$

$$n_{\text{No}} = \frac{5174}{7043} \times 288 \approx 212 \tag{17}$$

Dengan demikian, sampel penelitian terdiri dari 76 pelanggan mengalami *churn (Yes)* dan 212 pelanggan tidak mengalami *churn (No)* yang dipilih secara acak proporsional. Penggunaan rumus Cochran memastikan ukuran sampel cukup untuk mewakili populasi secara statistik, sedangkan metode *stratified random sampling* menjamin distribusi sampel seimbang antar kelompok. Kombinasi keduanya menghasilkan sampel yang representatif dan dapat meningkatkan keakuratan analisis regresi logistik biner terhadap faktor-faktor yang memengaruhi pelanggan mengalami churn.

3.3 Statistika Deskriptif

Analisis ini bertujuan untuk mengetahui bagaimana distribusi, frekuensi dan proporsi dari setiap variabel yang digunakan.

Tabel 2: Hasil Analisis Statistik Deskriptif

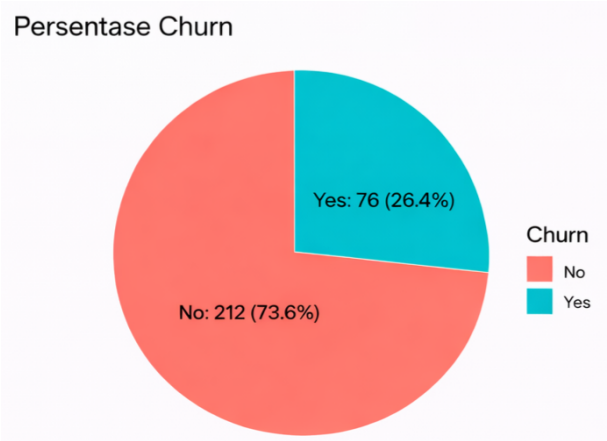
Variabel	N	Minimum	Maximum	Mean	Std. Deviasi
Churn (Y)	288	0	1	0.26	0.44
Layanan Telepon (X_1)	288	0	1	0.90	0.31
Layanan Internet (X_2)	288	0	1	0.44	0.50
Kontrak (X_3)	288	0	1	0.54	0.50

Berdasarkan hasil analisis statistik deskriptif yang terlihat pada Tabel 2, diketahui bahwa variabel dianalisis memiliki jumlah data sebanyak 288 pelanggan. Variabel *churn* (Y) rata-ratanya adalah 0.26 yang artinya sekitar 26% pelanggan dalam dataset tercatat melakukan *churn*, sedangkan 74% lainnya masih bertahan sebagai pelanggan. Nilai standar deviasi sebesar 0.44 mengindikasikan bahwa ada perbedaan nyata antara pelanggan yang keluar (*churn*) dan pelanggan yang tetap.

Selanjutnya, variabel layanan telepon (X_1) memiliki rata-rata 0.90 dan standar deviasi 0.31. Artinya, sebagian besar pelanggan yaitu sekitar 90% menggunakan layanan telepon, sedangkan hanya sebagian kecil yang tidak menggunakan layanan tersebut. Ini menunjukkan bahwa penggunaan layanan telepon sangat mendominasi di antara pelanggan. Di sisi lain, variabel layanan internet (X_2) memiliki rata-rata 0.44 dan standar deviasi 0.50, yang menunjukkan bahwa 44% pelanggan menggunakan layanan internet fiber optic, sementara 56% lainnya menggunakan jenis layanan internet yang berbeda atau tidak berlangganan internet sama sekali.

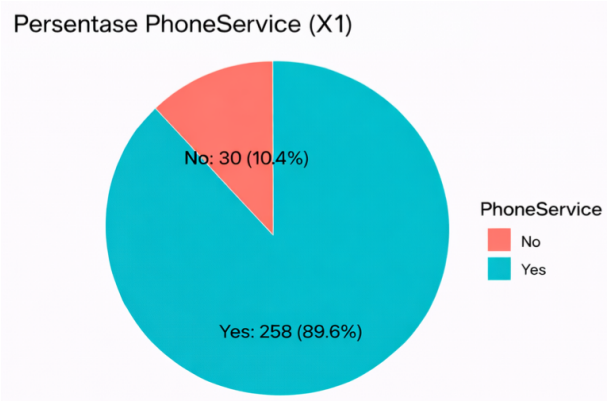
Selanjutnya, variabel kontrak (X_3) memiliki rata-rata 0.54 dan standar deviasi 0.50 yang menunjukkan bahwa sekitar 54% pelanggan memilih kontrak bulanan (*month-to-month*), sedangkan 46% pelanggan memilih kontrak jangka panjang seperti satu atau dua tahun. Secara keseluruhan, hasil ini menunjukkan bahwa sebagian besar pelanggan masih menggunakan layanan telepon, distribusi pelanggan berdasarkan jenis kontrak tergolong seimbang, dan penggunaan layanan internet fiber masih belum menjadi pilihan utama. Kombinasi karakteristik ini menjadi dasar awal untuk memahami faktor-faktor yang mungkin memengaruhi kecenderungan pelanggan untuk berhenti menggunakan layanan (*churn*).

Analisis deskripsi dari setiap variabel dapat dilihat pada diagram dibawah ini:



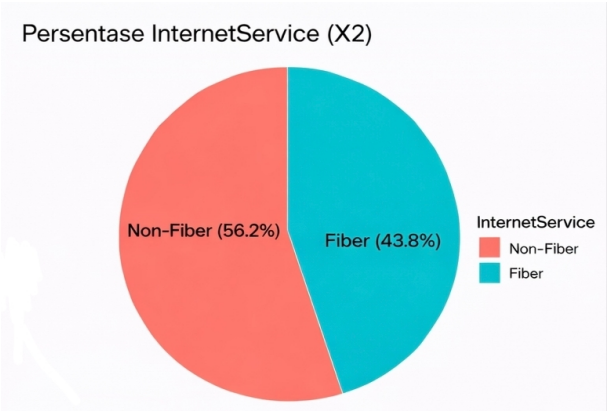
Gambar 1: Persentase variabel churn

Berdasarkan diagram pie pada Gambar 1, terlihat bahwa jumlah pelanggan yang tidak mengalami *churn* (*No*) mencapai 73.6% jauh lebih banyak dibandingkan dengan pelanggan yang mengalami *churn* (*Yes*) hanya sebesar 26.4%.



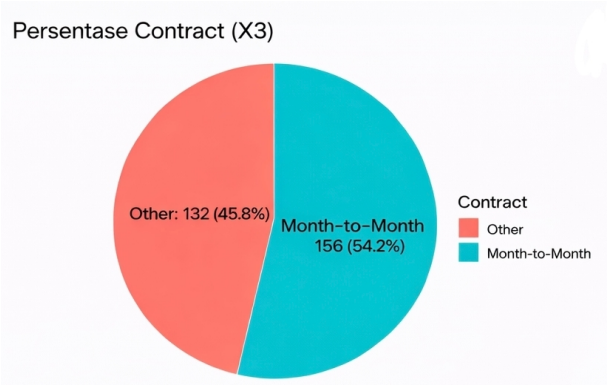
Gambar 2: Persentase variabel layanan telepon

Berdasarkan diagram pada Gambar 2, terlihat bahwa sebagian besar pelanggan menggunakan layanan telepon dengan persentase 89.6%, sedangkan hanya 10.4% pelanggan yang tidak menggunakan layanan tersebut. Penggunaan layanan telepon yang mendominasi menunjukkan bahwa akses telekomunikasi suara tetap menjadi kebutuhan utama bagi sebagian besar pelanggan.



Gambar 3: Persentase variabel layanan internet

Berdasarkan diagram pada Gambar 3, menampilkan bagaimana pelanggan menggunakan layanan internet. Terlihat 43.8% menggunakan layanan internet fiber optic, sedangkan 56.2% lainnya masih memilih layanan non-fiber. Angka ini menunjukkan bahwa meskipun layanan fiber optic semakin diinginkan dan menyediakan kualitas akses yang lebih baik, lebih dari separuh pelanggan tetap memilih layanan lain. Hal ini mungkin karena layanan tersebut lebih murah atau tersedia di wilayah tempat mereka tinggal.



Gambar 4: Persentase variabel kontrak

Berdasarkan diagram pada Gambar 4, terlihat bahwa jenis kontrak pelanggan terbagi secara

seimbang, dengan 54.2% pelanggan memilih kontrak bulanan (*month-to-month*) dan 45.8% memilih kontrak jangka panjang. Distribusi ini menunjukkan bahwa sebagian besar pelanggan memilih kontrak bulanan karena fleksibilitasnya, yang dalam teori bisa membuat pelanggan lebih mungkin untuk berpindah ke penyedia lain. Di sisi lain, hampir setengah pelanggan memilih kontrak jangka panjang, yang biasanya memberikan rasa aman dan kesepakatan yang lebih tetap.

3.4 Asumsi Multikolinearitas

Multikolinearitas merupakan kondisi dimana variabel prediktor saling berkorelasi tinggi, yang dapat memengaruhi stabilitas koefisien dalam model regresi logistik. Berdasarkan hasil perhitungan VIF pada model, diperoleh hasil sebagai berikut:

Tabel 3: Hasil estimasi Nilai VIF

Variabel	VIF
Layanan Telepon (X_1)	1,129715
Layanan Internet (X_2)	1,142078
Kontrak (X_3)	1,011633

Dari Tabel 3, nilai VIF untuk semua variabel prediktor berada di bawah 5, yaitu berkisar dari 1.01 sampai 1.14. Nilai VIF yang rendah ini menunjukkan bahwa tidak ada masalah multikolinearitas antar variabel prediktor dalam model tersebut. Hasil ini mengindikasikan bahwa model regresi logistik yang dibangun stabil dan koefisien prediktor dapat diinterpretasi dengan baik tanpa bias akibat multikolinearitas.

3.5 Uji Simultan

Untuk mengetahui apakah semua variabel prediktor secara bersamaan memiliki pengaruh yang signifikan terhadap variabel respon pada tingkat signifikansi tertentu dalam model regresi secara keseluruhan.

Hipotesis:

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ (Artinya, semua variabel independen tidak ada yang mempengaruhi variabel dependen).

H_1 : setidaknya satu $\beta_i \neq 0 \quad (i = 1, 2, 3)$ (Artinya, setidaknya ada satu variabel independen yang memengaruhi variabel dependen secara signifikan).

Dengan menggunakan Program RGui, dapat dilihat hasil dari uji *Likelihood Ratio Chi-Square* (G^2):

Tabel 4: Likelihood Ratio Chi-Square

Statistik	llh	llhNull	G^2	McFadden R^2	R^2_{ML}	R^2_{CU}
Nilai	-134.63007	-166.20059	63.14104	0.18995	0.19687	0.28753

Dengan tingkat signifikansi 0.05, diperoleh nilai statistik uji G^2 pada Tabel 4 yaitu 63.14104. Nilai ini dibandingkan dengan nilai kritis dari distribusi chi-kuadrat yaitu 7.815. Karena nilai G^2 lebih besar dari nilai kritis, maka kesimpulan yang diambil adalah menolak H_0. Artinya, secara bersamaan semua variabel independen yang dimasukkan ke dalam model mempunyai pengaruh signifikan terhadap variabel dependen.

3.6 Uji Parsial

Uji parsial dilakukan untuk menilai apakah variabel prediktor secara individu berpengaruh terhadap variabel respon. Adapun hasil masing-masing prediktornya antara lain:

Tabel 5: Hasil uji parsial (wald test) pada model regresi logistik

Variabel	Estimate	Std. Error	z value	Pr ($>\|z\|$)
Intercept	-3.0586	0.6323	-4.837	1.32e-06
Layanan Telepon (X_1)	0.1463	0.6204	0.236	0.813610
Layanan Internet (X_2)	1.1053	0.3272	3.377	0.000732
Kontrak (X_3)	1.9549	0.3751	5.212	1.87e-07

3.6.1 Uji Parsial dengan Variabel Prediktor Layanan Telepon

Hipotesis:

$H_0 : \beta_1 = 0$, artinya tidak terdapat pengaruh signifikan antara kepemilikan layanan telepon (*phone service*) terhadap kemungkinan pelanggan melakukan churn.

$H_1 : \beta_1 \neq 0$, artinya terdapat pengaruh signifikan antara kepemilikan layanan telepon (*phone service*) terhadap kemungkinan pelanggan melakukan churn.

$\alpha = 0.05$

Kriteria Keputusan: Tolak H_0 jika $|W| > Z_{(\alpha/2)} = 1.96$

Berdasarkan Tabel 5 terlihat bahwa nilai z-value atau Uji wald pada variabel layanan telepon yaitu $|0.236| < 1.96$ yang menyatakan bahwa variabel layanan telepon tidak memiliki pengaruh signifikan secara individu terhadap kemungkinan pelanggan melakukan *churn*, maka H_0 diterima.

3.6.2 Uji Parsial dengan Variabel Prediktor Layanan Internet

Hipotesis:

$H_0 : \beta_2 = 0$, artinya tidak terdapat pengaruh signifikan antara penggunaan layanan internet fiber optic terhadap kemungkinan pelanggan melakukan churn.

$H_1 : \beta_2 \neq 0$, artinya terdapat pengaruh signifikan antara penggunaan layanan internet fiber optic terhadap kemungkinan pelanggan melakukan churn.

$\alpha = 0.05$

Kriteria Keputusan: Tolak H_0 jika $|W| > Z_{(\alpha/2)} = 1.96$

Berdasarkan Tabel 5 terlihat bahwa nilai z-value atau Uji wald pada variabel layanan internet yaitu $|3.377| > 1.96$ maka H_0 ditolak yang artinya variabel layanan internet memiliki pengaruh signifikan secara individu terhadap kemungkinan pelanggan melakukan *churn*. Nilai koefisien positif menunjukkan bahwa pelanggan yang menggunakan layanan internet fiber optic memiliki peluang untuk berpindah ke penyedia lain lebih besar dibandingkan pelanggan yang tidak menggunakan layanan fiber optic.

3.6.3 Uji Parsial dengan Variabel Prediktor Kontrak

Hipotesis:

$H_0 : \beta_3 = 0$, artinya tidak ada pengaruh yang signifikan dari jenis kontrak terhadap kemungkinan pelanggan melakukan churn.

$H_1 : \beta_3 \neq 0$, artinya ada pengaruh yang signifikan dari jenis kontrak terhadap kemungkinan pelanggan melakukan churn.

$\alpha = 0.05$

Kriteria Keputusan: Tolak H_0 jika $|W| > Z_{(\alpha/2)} = 1.96$

Berdasarkan Tabel 5 terlihat bahwa nilai z-value atau Uji wald pada variabel kontrak yaitu $|5.212| > 1.96$ maka H_0 ditolak yang artinya variabel tipe kontrak memiliki pengaruh signifikan secara individu terhadap kemungkinan pelanggan melakukan churn. Koefisien bertanda positif menunjukkan bahwa pelanggan dengan kontrak *month-to-month* cenderung lebih mungkin untuk memutus kontraknya dibandingkan pelanggan yang memiliki kontrak dengan durasi yang lebih lama.

3.7 Uji Simultan Pada Variabel Signifikan

Hasil uji signifikansi menunjukkan bahwa variabel X_1 yaitu layanan telepon (*phone service*) tidak berpengaruh signifikan terhadap variabel dependen yaitu *churn*. Oleh sebab itu, variabel yang tidak signifikan dikeluarkan agar model regresi logistik biner menjadi lebih efisien dan akurat.

Untuk menguji apakah kedua variabel prediktor secara simultan memiliki pengaruh yang signifikan terhadap variabel respon pada suatu tingkat signifikansi tertentu dalam model regresi secara keseluruhan.

Hipotesis:

$H_0 : \beta_2 = \beta_3 = 0$ (Artinya, semua variabel independen tidak ada yang mempengaruhi variabel dependen).

H_1 : setidaknya satu $\beta_i \neq 0 \quad (i = 2, 3)$ (Artinya, setidaknya ada satu variabel independen yang memengaruhi variabel dependen secara signifikan).

Dengan menggunakan Program RGui, dapat dilihat hasil dari uji *Likelihood Ratio Chi-Square* (G^2):

Tabel 6: Uji Simultan untuk Variabel Signifikan

Statistik	llh	llhNull	G^2	McFadden R^2	R^2_{ML}	R^2_{CU}
Nilai	-134.65839	-166.20059	63.08441	0.18978	0.19671	0.28730

Dengan tingkat signifikansi sebesar $\alpha = 0.05$, diperoleh nilai statistik uji G^2 pada Tabel 6 yaitu 63.08441. Nilai ini dibandingkan dengan nilai kritis dari distribusi chi-kuadrat yaitu 5.991. Karena nilai G^2 lebih besar dari nilai kritis tersebut, maka keputusan yang diambil adalah menolak H_0. Artinya, secara simultan variabel-variabel independent berpengaruh signifikan terhadap variabel dependen.

3.8 Uji Parsial Pada Variabel Signifikan

Uji parsial dilakukan untuk mengetahui pengaruh setiap variabel prediktor secara parsial terhadap variabel respon. Hasil uji untuk setiap variabel prediktor menggunakan program RGui yaitu sebagai berikut:

Tabel 7: Uji parsial variabel signifikan menggunakan uji wald

Variabel	Estimate	Std. Error	z value	Pr ($>\|z\|$)
Intercept	-2.9382	0.3664	-8.019	1.07e-15
Layanan Internet (X_2)	1.1323	0.3079	3.678	0.000235
Kontrak (X_3)	1.9535	0.3750	5.209	1.90e-07

3.8.1 Uji Parsial dengan Variabel Prediktor Layanan Internet

Hipotesis:

$H_0 : \beta_2 = 0$, artinya tidak terdapat pengaruh signifikan antara penggunaan layanan internet fiber optic terhadap kemungkinan pelanggan melakukan churn.

$H_1 : \beta_2 \neq 0$, artinya terdapat pengaruh signifikan antara penggunaan layanan internet fiber optic terhadap kemungkinan pelanggan melakukan churn.

$\alpha = 0.05$

Kriteria Keputusan: Tolak H_0 jika $|W| > Z_{(\alpha/2)} = 1.96$.

Berdasarkan Tabel 7 terlihat nilai z-value atau Uji wald pada variabel layanan internet yaitu $|3.678| > 1.96$ maka H_0 ditolak yang artinya variabel layanan internet memiliki pengaruh signifikan secara individu terhadap kemungkinan pelanggan melakukan *churn*. Nilai koefisien positif menunjukkan bahwa pelanggan yang menggunakan layanan internet fiber optic memiliki

peluang untuk berpindah ke penyedia lain lebih besar dibandingkan pelanggan yang tidak menggunakan layanan fiber optic.

3.8.2 Uji Parsial dengan Variabel Prediktor Kontrak

Hipotesis:

$H_0 : \beta_3 = 0$„ artinya tidak ada pengaruh yang signifikan dari jenis kontrak terhadap kemungkinan pelanggan melakukan *churn*.

$H_1 : \beta_3 \neq 0$, artinya ada pengaruh yang signifikan dari jenis kontrak terhadap kemungkinan pelanggan melakukan churn.

$\alpha = 0.05$

Kriteria Keputusan: Tolak H_0 jika $|W| > Z_{(\alpha/2)} = 1.96$.

Berdasarkan Tabel 7 terlihat nilai z-value atau Uji wald pada variabel kontrak yaitu $|5.209| > 1.96$ maka H_0 ditolak yang artinya variabel tipe kontrak memiliki pengaruh signifikan secara individu terhadap kemungkinan pelanggan melakukan *churn*. Koefisien bertanda positif menunjukkan bahwa pelanggan dengan kontrak *month-to-month* cenderung lebih mungkin untuk memutus kontraknya dibandingkan pelanggan yang memiliki kontrak dengan durasi yang lebih lama.

3.9 Uji Kecocokan Model

Uji kecocokan model digunakan untuk memeriksa apakah model yang digunakan cocok dengan data yang tersedia atau tidak.

Hipotesis yang digunakan yaitu:

H_0 : model yang dihipotesiskan cocok dengan data

H_1 : model yang dihipotesiskan tidak cocok dengan data

Kriteria pengambilan keputusan: jika nilai p-value > 0.05 maka terima H_0.

Dengan menggunakan program RGui, diperoleh hasil seperti pada table dibawah:

Tabel 8: Uji Hosmer dan Lemeshow

Step	Chi-squared	df	p-value
1	0.073017	1	0.787

Dari Tabel 8, diperoleh nilai p-value > 0.05 , maka keputusan terima H_0. Artinya, model yang dihipotesiskan fit dengan data. Artinya, model mampu merepresentasikan hubungan antara variabel independent dan dependen dengan baik serta layak digunakan.

3.10 Odds Ratio dan Interpretasi Model Regresi Logistik

Berdasarkan (4) bentuk model regresi logistik multivariat dapat dituliskan sebagai berikut:

$$\hat{\pi}(x) = \frac{\exp(-2.9382 + 1.1323x_2 + 1.9535x_3)}{1 + \exp(-2.9382 + 1.1323x_2 + 1.9535x_3)}$$

Berdasarkan (10) model regresi logistik ditulis dalam bentuk logit, maka menjadi:

$$g(x) = -2.9382 + 1.1323\, x_2 + 1.9535\, x_3$$

Hasil estimasi menunjukkan bahwa variabel layanan internet memiliki koefisien positif dengan odds ratio sebesar $e^{1.1323} = 3.10$, yang berarti pelanggan yang menggunakan layanan internet berbasis fiber optic memiliki peluang untuk berhenti menggunakan layanan (*churn*) sebesar 3.10 kali lebih besar dibandingkan pelanggan yang menggunakan layanan non-fiber. Sebaliknya, variabel kontrak memiliki koefisien positif sebesar 1.9535 dengan odds ratio $e^{1.9535} = 7.05$, berarti

pelanggan yang memiliki kontrak bulanan *(month-to-month)* memiliki peluang *churn* sebesar 7.05 kali lebih tinggi dibandingkan pelanggan dengan kontrak jangka panjang. Dengan demikian, jenis kontrak dianggap sebagai faktor yang paling berpengaruh dalam meningkatkan peluang churn dibandingkan dengan jenis layanan internet.

Apabila kedua variabel bernilai 0 artinya pelanggan menggunakan internet non-fiber dan masih memiliki kontrak jangka panjang, maka nilai logit adalah –2.9382, yang berarti risiko churn sekitar 5%. Jika pelanggan menggunakan layanan fiber optic ($X_2(1)$) namun tetap memiliki kontrak jangka panjang ($X_3(0)$), nilai logit berubah menjadi –1,8059 dan risiko churn meningkat menjadi sekitar 14.2%. Hal ini menunjukkan bahwa penggunaan layanan fiber optic justru meningkatkan kemungkinan pelanggan mengakhiri kontrak, meskipun mereka masih dalam kontrak jangka panjang.

Selanjutnya, apabila pelanggan memiliki kontrak bulanan ($X_3(1)$), tetapi menggunakan layanan non-fiber ($X_2(0)$), nilai logit meningkat menjadi -0.9847 sehingga probabilitas churn mencapai 27.1%. Artinya, fleksibilitas kontrak bulanan memberi peluang yang lebih besar bagi pelanggan untuk berhenti menggunakan layanan, meskipun mereka tidak memakai layanan fiber optic. Sementara itu, pada kondisi di mana pelanggan menggunakan fiber optic sekaligus memiliki kontrak bulanan ($X_2 = 1, X_3 = 1$), nilai logit menjadi 0.1476 dengan probabilitas *churn* meningkat tajam menjadi 53.7%, yang merupakan kondisi dengan tingkat *churn* tertinggi.

Secara keseluruhan, hasil ini menunjukkan bahwa variabel Kontrak (X_3) mempunyai pengaruh yang lebih besar terhadap kemungkinan pelanggan berhenti (*churn*) dibandingkan variabel Layanan Internet (X_2). Pelanggan yang memiliki kontrak bulanan (*month-to-month*) cenderung lebih mudah berhenti, sementara pelanggan yang menggunakan layanan fiber optic juga memiliki risiko *churn* yang lebih tinggi, meskipun risikonya tidak sebesar pelanggan dengan tipe kontrak. Temuan ini menunjukkan bahwa perlu adanya strategi penahanan pelanggan yang lebih intensif terhadap kelompok pelanggan dengan kontrak bulanan dan pengguna layanan fiber optic, karena kedua kelompok tersebut termasuk dalam segmen pelanggan dengan risiko churn terbesar.

Selain interpretasi odds ratio, kinerja model juga dinilai menggunakan beberapa metrik klasifikasi yang lebih informatif untuk kasus churn. Berdasarkan tabel klasifikasi, diperoleh tingkat akurasi model sebesar 75,35%, yang menunjukkan bahwa model memiliki kemampuan sedang dalam membedakan pelanggan yang mengalami churn dan yang tidak mengalami churn. Hasil ini didukung oleh Tabel 9, yang menggambarkan kinerja model dalam mengklasifikasikan masing-masing kelas pelanggan.

Tabel 9: Tabel Crosstab

Actual	Predicted	No	Yes	Percentage Correct (%)
Churn	No	170	42	80.2
	Yes	29	47	61.8
Overall Percentage				75.35

Berdasarkan Tabel 9, nilai spesifisitas model untuk kelas *non-churn* mencapai 80,2%, yang menunjukkan bahwa model cukup baik dalam mengidentifikasi pelanggan yang tetap berlangganan. Sementara itu, nilai sensitivitas (recall) untuk kelas churn sebesar 61,8%, yang menunjukkan bahwa masih terdapat sebagian pelanggan churn yang belum berhasil teridentifikasi oleh model. Perbedaan nilai sensitivitas dan spesifisitas ini mengindikasikan bahwa model cenderung lebih akurat dalam mengklasifikasikan pelanggan *non-churn* dibandingkan pelanggan churn, yang merupakan karakteristik umum pada data churn dengan distribusi kelas yang tidak seimbang.

Perbedaan ini menunjukkan bahwa model lebih baik dalam mengidentifikasi pelanggan yang tetap berlangganan dibandingkan pelanggan yang berpotensi *churn*. Secara keseluruhan, model regresi logistik yang telah dibuat mampu mengklasifikasikan dengan tepat sebanyak 75.35% dari

total pengamatan. Sehingga besarnya missklasifikasi dapat dihitung sebagai berikut:

$$\text{Miss Klasifikasi} = \frac{42 + 29}{170 + 42 + 29 + 47} = \frac{71}{288} \approx 24.65\%$$

Nilai ini menunjukkan bahwa meskipun model sudah mampu memberikan hasil klasifikasi yang cukup bagus, masih ada kemungkinan untuk meningkatkan kemampuan model dalam mendeteksi pelanggan yang akan berpindah. Keterbatasan tersebut mungkin disebabkan oleh jumlah variabel prediktor yang digunakan, yang belum cukup lengkap hanya mengandalkan variabel layanan inti.

Secara konseptual, evaluasi kemampuan model juga bisa dilakukan dengan melihat kurva ROC dan nilai Area Under the Curve (AUC) untuk menilai seberapa baik model dalam membedakan pelanggan yang berpotensi churn dan yang tidak pada berbagai tingkat keputusan. Meskipun dalam penelitian ini analisis ROC–AUC tidak ditampilkan secara eksplisit, metrik ini bisa digunakan di penelitian berikutnya untuk memberikan gambaran lebih lengkap mengenai kinerja model, terutama pada data churn yang tidak seimbang.

Dari sisi praktis, temuan ini menunjukkan bahwa pelanggan yang menggunakan layanan internet fiber memiliki risiko *churn* yang cukup tinggi. Hal ini menunjukkan bahwa pelanggan fiber memiliki ekspektasi terhadap kualitas layanan yang lebih tinggi dan lebih peka terhadap gangguan jaringan, harga, maupun kualitas layanan setelah pembelian. Oleh karena itu, perusahaan perlu fokus pada strategi retensi untuk segmen ini dengan meningkatkan stabilitas jaringan, mempercepat respons terhadap keluhan, serta memberikan program loyalitas yang sesuai dengan karakteristik pelanggan fiber. Selain itu, tingginya risiko churn pada pelanggan dengan kontrak bulanan menunjukkan bahwa fleksibilitas kontrak memberikan kesempatan lebih besar bagi pelanggan untuk berpindah ke penyedia lain. Implikasi praktisnya adalah perlu adanya strategi retensi yang lebih aktif bagi pelanggan kontrak bulanan, seperti penawaran insentif untuk memperpanjang kontrak, penawaran paket layanan yang lebih lengkap, atau program loyalitas berdasarkan penggunaan, agar bisa mengurangi potensi churn di segmen pelanggan tersebut.

4 Kesimpulan

Penelitian ini menunjukkan bahwa jenis kontrak dan layanan internet memainkan peran penting dalam menentukan risiko pelanggan berhenti menggunakan layanan. Dari variabel yang dianalisis, pelanggan yang mengikuti kontrak bulanan memiliki risiko churn yang paling tinggi, yaitu sekitar 7,05 kali lebih besar dibandingkan pelanggan dengan kontrak jangka panjang. Hasil ini menunjukkan bahwa kontrak yang lebih fleksibel memungkinkan pelanggan lebih mudah beralih ke penyedia lain. Selain itu, pelanggan yang menggunakan layanan internet fiber juga memiliki risiko churn yang tinggi, sekitar 3,10 kali lebih besar dibandingkan pelanggan non-fiber, yang menunjukkan bahwa kelompok ini memiliki standar kualitas layanan yang lebih tinggi.

Dari sisi penerapan, hasil penelitian ini menekankan perlunya strategi retensi yang lebih efektif terutama untuk pelanggan kontrak bulanan. Perusahaan disarankan untuk mendorong pelanggan kontrak bulanan untuk beralih ke kontrak jangka panjang dengan menawarkan insentif tambahan, paket layanan yang lebih menarik, serta program loyalitas berdasarkan penggunaan layanan. Untuk pelanggan fiber, perusahaan juga perlu meningkatkan kualitas jaringan, mengatasi gangguan dengan cepat, serta memperkuat layanan setelah pembelian agar kepuasan pelanggan tetap terjaga dan risiko churn dapat diminimalkan.

Penelitian ini memiliki beberapa keterbatasan, seperti penggunaan variabel prediktor yang hanya mencakup layanan inti, belum mempertimbangkan efek interaksi antar variabel, serta evaluasi kinerja model yang masih menggunakan metrik klasifikasi sederhana tanpa analisis ROC–AUC. Untuk penelitian selanjutnya, disarankan untuk menambahkan variabel perilaku pelanggan, menganalisis hubungan antar variabel, serta menggunakan metrik evaluasi dan pendekatan pemodelan yang lebih lengkap agar kemampuan prediksi dan interpretasi model churn menjadi lebih baik.

Pernyataan Kontribusi Penulis (CRediT)

Zahwa Rifsya Pangest: Konseptualisasi, Metodologi, Perangkat Lunak, Validasi, Analisis Formal, Investigasi, Kurasi Data, Visualisasi, Penulisan Draft Awal. **Corry Sormin**: Supervisi, Penulisan Telaah dan Penyuntingan, Administrasi Proyek.

Deklarasi Penggunaan AI atau Teknologi Berbasis AI

Model ChatGPT versi 5.1 digunakan dengan pembatasan untuk membantu menyusun draf awal kalimat, memperbaiki struktur bahasa, melakukan penyuntingan redaksional, memperoleh referensi sintaks R, memberikan petunjuk dalam melakukan analisis, serta membantu mengatasi kesalahan yang muncul selama proses pengolahan data.

Deklarasi Konflik Kepentingan

Penulis menyatakan bahwa penelitian ini dilakukan sebagai bagian dari pemenuhan persyaratan program MBKM Studi Independen dari Universitas.

Pendanaan dan Ucapan Terima Kasih

Penelitian ini menerima pendanaan oleh dosen pembimbing untuk mendukung proses penyelesaian dan penerbitan artikel. Selain itu, penelitian ini tidak mendapat dana dari lembaga di luar institusi. Penulis ingin menyampaikan terima kasih kepada dosen pembimbing yang telah memberikan arahan, dukungan ilmiah serta bantuan dana untuk publikasi sehingga penelitian dan penulisan artikel ini dapat diselesaikan dengan baik.

Ketersediaan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang berasal dari dataset publik yang berkaitan dengan churn pelanggan [1] di sektor telekomunikasi dan dapat diakses secara terbuka untuk kepentingan penelitian dan pembelajaran. Sebelum dilakukan analisis, data telah melewati tahap pra-pemrosesan agar sesuai dengan kebutuhan penerapan regresi logistik biner. Seluruh data yang dianalisis tidak mengandung informasi identitas pelanggan, sehingga penggunaannya telah memenuhi prinsip etika penelitian dan ketentuan perlindungan data.

Daftar Pustaka

[1] S. D. Damanik and M. I. Jambak, "Klasifikasi customer churn pada telekomunikasi industri untuk retensi pelanggan menggunakan algoritma c4.5," *J. Sist. Inf. Bisnis*, vol. 3, no. 6, pp. 1303–1309, 2023. DOI: 10.30865/klik.v3i6.829

[2] Y. Yudiana, A. Yulia, and N. Khofifah, "Prediksi customer churn menggunakan metode crisp-dm pada industri telekomunikasi sebagai implementasi mempertahankan pelanggan," *Indones. J. Islam. Econ. Bus.*, vol. 8, no. 1, pp. 1–20, 2023.

[3] A. Adiansya and Z. Abidin, "The implementation of a logistic regression algorithm and gradient boosting classifier for predicting telco customer churn," *Pixel: Jurnal Ilmiah Komputer Grafis*, vol. 17, no. 1, pp. 168–178, 2024.

[1] https://www.kaggle.com/datasets/blastchar/telco-customer-churn?resource=download

[4] A. Nurtriana, D. D. Rachmawati, M. Artiyasa, D. Syahrudin, and Z. Sidiq, "Churn prediction analysis of telecom customers using svm, random forest and logistic regression models using orange data mining tools," *International Conference on Computer Science Electronics and Information*, vol. 02012, 2024.

[5] S. Antoh, R. Herteno, I. Budiman, D. Kartini, and M. I. Mazdadi, "Prediksi churn pelanggan telekomunikasi dengan optimalisasi seleksi fitur dan tuning hyperparameter pada algoritma klasifikasi c4.5," *Jurnal Sistem Informasi Bisnis*, vol. 15, no. 1, pp. 60–67, 2025. DOI: 10.14710/vol15iss1pp60-67

[6] L. N. Wakhidah, A. K. Zyen, and B. B. Wahono, "Evaluation of telecommunication customer churn classification with smote using random forest and xgboost algorithms," *Journal of Applied Informatics and Computing*, vol. 9, no. 1, pp. 89–95, 2025.

[7] S. G. Putu, "Menentukan populasi dan sampel; pendekatan metodologi penelitian kuantitatif dan kualitatif," *J. Ilm. Profesi Pendidik.*, vol. 9, pp. 2721–2731, 2024.

[8] B. Sumargo, *Teknik Sampling*. Unj Press, 2020.

[9] E. Roflin, F. Riana, E. Munarsih, and I. A. Liberty, *Regresi Logistik Biner dan Multinomial*. Penerbit NEM, 2023.

[10] D. Kartikasari, "Analisis faktor-faktor yang mempengaruhi level polusi udara dengan metode regresi logistik biner," *MATHunesa J. Ilm. Mat.*, vol. 8, no. 1, pp. 55–59, 2020.

[11] A. A. Rahmadani et al., "Analisis regresi logistik biner untuk memprediksi faktor-faktor internal yang memengaruhi keharmonisan rumah tangga menurut provinsi di indonesia pada tahun 2021," *J. FMIPA Unmul*, vol. 3, no. 1, pp. 116–127, 2023. Available online.

[12] I. Susanti and F. Saumi, "Penerapan metode analisis regresi linear berganda untuk mengatasi masalah multikolinearitas pada kasus indeks pembangunan manusia (ipm) di kabupaten aceh tamiang," *Gamma-Pi J. Mat. dan Terap.*, vol. 4, no. 2, pp. 38–42, 2022.

[13] M. A. Santika and Y. Karyana, "Analisis regresi logistik biner dengan efek interaksi untuk memodelkan angka fertilitas total di jawa barat," *Bandung Conf. Ser. Stat.*, vol. 2, no. 2, pp. 142–151, 2022. DOI: 10.29313/bcss.v2i2.3555

[14] S. A. R. Manaf, Erfiani, Indahwati, A. Fitrianto, and R. Amelia, "Faktor-faktor yang memengaruhi permasalahan stunting di jawa barat menggunakan regresi logistik biner," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 15, no. 2, pp. 265–274, 2022. DOI: 10.36456/jstat.vol15.no2.a5654

[15] R. F. Kaban and D. A. Purnawarman, "How gender, region, and lifestyle effect the decision to islamic financial literacy of millennial generation in greater jakarta," *Indik. J. Ilm. Manaj. dan Bisnis*, vol. 8, no. 3, p. 14, 2024. DOI: 10.22441/indikator.v8i3.28104

[16] W. Alwi, E. Ermawati, and S. Husain, "Analisis regresi logistik biner untuk memprediksi kepuasan pengunjung pada rumah sakit umum daerah majene," *J. MSA (Mat. dan Stat. serta Apl.)*, vol. 6, no. 1, p. 20, 2018. DOI: 10.24252/msa.v6i1.4783

[17] D. W. H. Jr, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. John Wiley & Sons, 2013.

[18] A. I. Sofiyat, A. Tjalla, and Mahdiyah, "Pemodelan regresi logistik biner terhadap penerimaan pegawai di pt xyz jakarta," *Mat. Sains*, vol. 1, no. 1, pp. 1–11, 2023.