

# *Prediksi Pelanggan Churn pada Perusahaan Telekomunikasi Menggunakan Logistic Regression di RapidMiner*

Andri Tri Hartanto

Program Studi : Sistem Informasi, Universitas Darwan Ali

Sistem Informasi, Fakultas: Ilmu Komputer

Email : [2357201002312@unda.ac.id](mailto:2357201002312@unda.ac.id)

**ABSTRACT**— This study aims to predict customer churn in the telecommunications sector using the Logistic Regression algorithm in RapidMiner. The dataset, sourced from Kaggle, comprises 7043 customer records with attributes such as contract type, monthly charges, tenure, and payment method. Initial exploratory analysis revealed a significant imbalance between churn and non-churn customer classes in the raw data. The research process involved preprocessing, data balancing using SMOTE, and modeling with Logistic Regression. The application of SMOTE successfully addressed the class imbalance issue, thereby enhancing the model's sensitivity to at-risk churn customers. The model achieved an accuracy of 90.72%, a recall of 85.95% for churn customers, and an AUC of 0.962. An AUC value close to 1 indicates the model's excellent capability in distinguishing between churn and non-churn customers. The most influential variables were contract type and monthly charges based on regression coefficient analysis. This finding confirms that customers with month-to-month contracts and higher charges exhibit a greater propensity to churn. The results provide crucial insights for customer retention strategies. Consequently, telecom operators can optimize loyalty programs and service package adjustments based on customer risk profiles.

**Keywords**— churn, logistic regression, rapidminer, SMOTE, telco.

**ABSTRAK**— Penelitian ini bertujuan memprediksi pelanggan churn di sektor telekomunikasi menggunakan algoritma Logistic Regression di RapidMiner. Dataset berasal dari Kaggle dan mencakup 7043 data pelanggan dengan atribut seperti jenis kontrak, biaya bulanan, lama berlangganan (tenure), dan metode pembayaran. Analisis eksploratori awal mengungkapkan ketidakseimbangan signifikan antara kelas pelanggan churn dan non-churn dalam data mentah. Proses penelitian melibatkan preprocessing, penyeimbangan data menggunakan SMOTE, dan pemodelan menggunakan Logistic Regression. Penerapan SMOTE berhasil mengatasi masalah ketidakseimbangan kelas sehingga meningkatkan sensitivitas model terhadap pelanggan berisiko churn. Model mencapai akurasi sebesar 90,72%, recall sebesar 85,95% untuk pelanggan churn, dan AUC sebesar 0,962. Nilai AUC yang mendekati 1 menunjukkan kemampuan model yang sangat baik dalam membedakan pelanggan churn dan non-churn. Variabel yang paling berpengaruh adalah jenis kontrak dan biaya bulanan berdasarkan analisis koefisien regresi. Temuan ini mengonfirmasi bahwa pelanggan dengan kontrak bulanan (month-to-month) dan biaya tinggi memiliki kecenderungan churn lebih besar. Hasil ini memberikan wawasan penting dalam strategi retensi pelanggan. Implikasinya, operator telekomunikasi dapat mengoptimalkan program loyalitas dan penyesuaian paket layanan berdasarkan profil risiko pelanggan.

**Kata kunci**— churn, logistic regression, rapidminer, SMOTE, telco.

## I. PENDAHULUAN

Churn pelanggan merupakan tantangan utama dalam industri telekomunikasi modern. Perusahaan di sektor ini menghadapi tingkat perputaran pelanggan yang tinggi sebagai akibat dari meningkatnya persaingan pasar, tersedianya banyak alternatif penyedia layanan, serta rendahnya biaya dan hambatan untuk berpindah layanan [1][11][13]. Selain dampak langsung terhadap pendapatan, churn juga berimplikasi pada citra merek dan loyalitas pelanggan. Tingkat churn yang tinggi sering kali mencerminkan ketidakpuasan layanan dan berdampak negatif pada reputasi perusahaan [7][12]. Perusahaan yang tidak mampu mengelola churn berisiko kehilangan keunggulan kompetitif dalam jangka panjang karena gagal membangun hubungan berkelanjutan dengan pelanggan. Oleh karena itu, mengelola churn bukan hanya upaya teknis dalam pengolahan data, melainkan strategi

bisnis yang bersifat holistik dan berorientasi jangka panjang.

Dalam konteks bisnis yang kompetitif, mempertahankan pelanggan menjadi prioritas karena akuisisi pelanggan baru terbukti memerlukan biaya yang jauh lebih besar, yakni sekitar 5–25 kali lipat dibandingkan mempertahankan pelanggan yang sudah ada [1]. Selain itu, kehilangan pelanggan juga berarti hilangnya potensi keuntungan jangka panjang yang biasa dikenal dengan konsep *customer lifetime value* (CLV). Dengan demikian, kemampuan untuk secara proaktif mengidentifikasi pelanggan yang berisiko berhenti sangat penting untuk memastikan keberlangsungan dan profitabilitas perusahaan. Melalui identifikasi dini terhadap pelanggan yang cenderung churn, perusahaan dapat menyusun strategi intervensi dan alokasi sumber daya secara lebih terfokus dan efisien [6].

Seiring berkembangnya teknologi dan ketersediaan data pelanggan yang semakin melimpah, pendekatan prediktif berbasis *machine learning* menjadi alternatif yang efektif dalam memitigasi churn. Model prediksi yang dibangun menggunakan pembelajaran mesin terbukti mampu meningkatkan akurasi identifikasi pelanggan berisiko hingga 30% dibanding pendekatan konvensional berbasis aturan atau intuisi manual [2][3][4]. Selain lebih presisi, pendekatan ini juga memungkinkan perusahaan menyesuaikan strategi bisnis secara *real time* berdasarkan hasil analisis data yang terus berkembang. Dengan kemampuan belajar dari pola data historis, model prediktif tidak hanya reaktif, tetapi juga proaktif dalam mendeteksi potensi churn sebelum terjadi. Oleh karena itu, teknologi ini menjadi alat penting dalam mengelola loyalitas pelanggan di era digital yang dinamis.

Salah satu tantangan utama yang sering dihadapi dalam analisis churn adalah ketidakseimbangan data, di mana jumlah pelanggan churn jauh lebih sedikit dibanding pelanggan yang bertahan. Ketidakseimbangan ini dapat mengganggu proses pembelajaran model karena algoritma cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas yang lebih penting [4][13]. Model prediksi yang efektif dapat digunakan untuk mengembangkan berbagai program retensi pelanggan yang lebih personal dan tertarget, seperti pemberian insentif, layanan khusus, atau rekomendasi produk yang relevan dengan kebutuhan pelanggan [6][9]. Strategi ini sejalan dengan prinsip *data-driven decision making*, yang menjadi kunci dalam pengambilan keputusan bisnis modern [13].

Di antara berbagai algoritma yang tersedia, Logistic Regression menjadi pilihan yang banyak diadopsi karena memiliki struktur sederhana, hasil yang mudah diinterpretasi, serta efisiensi tinggi dalam mengklasifikasikan masalah biner seperti churn [1][5]. Kelebihan lainnya adalah kemampuannya untuk mengukur seberapa besar kemungkinan seorang pelanggan akan churn dalam bentuk probabilitas, bukan hanya klasifikasi keras. Hal ini sangat berguna dalam penetapan prioritas intervensi bisnis karena memungkinkan penyusunan strategi berdasarkan tingkat risiko. Lebih lanjut, penelitian ini juga mengedepankan aspek interpretabilitas model dalam konteks penerapan bisnis, di mana Logistic Regression mendukung kebutuhan ini karena tidak hanya menghasilkan label klasifikasi, tetapi juga menyediakan informasi tentang variabel mana yang paling berkontribusi terhadap churn.

Untuk mengatasi ketidakseimbangan data, penelitian ini mengadopsi teknik SMOTE (Synthetic Minority Over-sampling Technique) guna meningkatkan proporsi data kelas minoritas dengan cara mensintesis contoh baru berdasarkan tetangga terdekat. Dengan metode ini, distribusi data menjadi lebih seimbang, sehingga model dapat lebih peka dalam mendeteksi pelanggan yang berisiko churn [5][10]. Dalam penelitian ini, penulis menggunakan RapidMiner sebagai *platform* pengolahan data dan pemodelan karena kemudahannya serta kemampuan visualisasi proses kerja yang sangat

mendukung alur *data science end-to-end* [8]. Transparansi dari Logistic Regression, ditambah dengan kemudahan penggunaan RapidMiner dan penanganan ketidakseimbangan data melalui SMOTE, menjembatani komunikasi antara tim teknis dan manajemen serta memastikan hasil model dapat diterima dan diimplementasikan secara tepat [6].

Penelitian ini juga berkontribusi pada aspek praktikalitas, di mana pendekatan yang digunakan dapat dengan mudah direplikasi oleh perusahaan lain, termasuk organisasi kecil dan menengah. Dengan memanfaatkan RapidMiner, perusahaan tidak perlu berinvestasi besar dalam pengembangan sistem internal atau sumber daya TI yang kompleks [8]. Proses yang ditawarkan bersifat efisien, cepat, dan ramah pengguna, sehingga dapat diterapkan bahkan oleh tim tanpa latar belakang pemrograman [7]. Hal ini membuka peluang lebih luas bagi sektor industri yang ingin mengadopsi pendekatan analitik churn secara strategis.

Secara keseluruhan, penelitian ini menegaskan bahwa prediksi churn pelanggan bukan hanya sebuah analisis statistik semata, melainkan bagian dari strategi proaktif untuk meningkatkan loyalitas, mempertahankan basis pelanggan, dan memaksimalkan nilai bisnis jangka panjang. Melalui pendekatan yang tepat dan teknologi yang mendukung, perusahaan dapat mengambil langkah antisipatif dalam menghadapi risiko churn serta membangun fondasi hubungan pelanggan yang lebih kuat dan berkelanjutan. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan kontribusi nyata baik dari sisi teknis analitik maupun dalam pengambilan keputusan strategis bisnis.

## II. METODOLOGI PENELITIAN

### 1. Dataset dan Preprocessing

Dataset yang digunakan dalam penelitian ini diperoleh dari sumber terbuka [6], yaitu dataset Telco Customer Churn yang terdiri dari 7.043 entri data pelanggan dan 21 atribut yang mencakup informasi demografis serta layanan yang digunakan pelanggan. Beberapa atribut penting antara lain *gender*, *SeniorCitizen*, *Partner*, *Dependents*, *tenure*, *PhoneService*, *InternetService*, *Contract*, *PaymentMethod*, hingga *Churn* sebagai target klasifikasi.

Langkah pertama dalam tahap *preprocessing* adalah pembersihan data. Terdapat nilai kosong (*missing value*) pada atribut *TotalCharges*, yang kemudian diisi dengan nilai median dari atribut tersebut untuk menjaga konsistensi distribusi data dan menghindari bias yang mungkin muncul jika diisi dengan nilai rata-rata.

Proses selanjutnya adalah mengubah atribut kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma Logistic Regression. Transformasi ini dilakukan menggunakan metode Nominal to Numerical, yaitu mengubah nilai-nilai kategori menjadi representasi angka dengan teknik seperti *label encoding* atau *one-hot encoding*, tergantung pada jenis data kategorikal. Terakhir, atribut *Churn* ditetapkan sebagai label atau

target yang akan diprediksi oleh model, di mana nilai "Yes" dikonversi menjadi 1 dan "No" menjadi 0.

## 2. Penyeimbangan Data

Salah satu tantangan utama dalam dataset Telco Customer Churn adalah ketidakseimbangan kelas (imbalanced class distribution). Hanya sekitar 26% pelanggan dalam dataset yang termasuk dalam kategori *churn*, sedangkan sisanya (74%) merupakan pelanggan yang tetap menggunakan layanan. Ketidakseimbangan ini dapat menyebabkan model bias terhadap kelas mayoritas (*non-churn*), sehingga performa model dalam mendeteksi pelanggan yang akan berhenti berlangganan menjadi rendah.

Untuk mengatasi masalah ini, digunakan teknik SMOTE (Synthetic Minority Over-sampling Technique), yaitu metode *oversampling* yang menghasilkan data sintesis pada kelas minoritas. SMOTE bekerja dengan membuat data baru berdasarkan interpolasi antara data minoritas yang ada dengan tetangga terdekatnya. Dalam penelitian ini, parameter SMOTE disesuaikan dengan percentage sebesar 200% dan jumlah *nearest neighbors* ( $k$ ) sebesar 5. Dengan demikian, jumlah data pada kelas *churn* ditingkatkan agar sebanding dengan jumlah data pada kelas *non-churn*, sehingga model dapat belajar secara seimbang terhadap kedua kelas.

## 3. Pemodelan dan Evaluasi

Pemodelan prediksi *churn* dilakukan dengan menggunakan algoritma Logistic Regression, yang merupakan metode klasifikasi biner berbasis regresi linier untuk memperkirakan probabilitas keanggotaan suatu kelas. Logistic Regression dipilih karena model ini mudah diinterpretasikan, efisien, dan banyak digunakan dalam studi prediksi *churn*.

Agar hasil pemodelan lebih akurat dan dapat digeneralisasi, digunakan metode validasi silang 5-fold cross-validation. Teknik ini membagi data menjadi lima subset yang sama besar; empat bagian digunakan untuk pelatihan (*training*), dan satu bagian untuk pengujian (*testing*), dilakukan secara bergantian sebanyak lima kali. Hasil dari tiap iterasi kemudian dirata-rata untuk mendapatkan performa keseluruhan model.

Evaluasi performa model dilakukan dengan menggunakan beberapa metrik penting, yaitu:

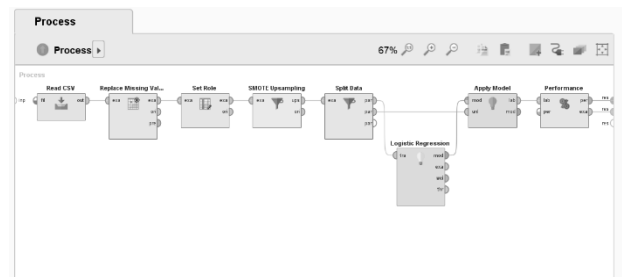
- Akurasi: menunjukkan proporsi prediksi yang benar terhadap seluruh data.
- Recall: mengukur kemampuan model dalam mendeteksi pelanggan yang benar-benar *churn* (*true positive rate*).
- Precision: mengukur proporsi prediksi *churn* yang benar-benar tepat.
- AUC (Area Under Curve): menunjukkan kemampuan model dalam membedakan antara kelas *churn* dan *non-churn* secara menyeluruh.

Dengan kombinasi *preprocessing* yang tepat, penyeimbangan data menggunakan SMOTE, serta pemodelan dan evaluasi yang sistematis, diharapkan model Logistic Regression yang dibangun mampu

memberikan hasil yang akurat dan reliabel dalam memprediksi *churn* pelanggan.

## III. DESAIN, HASIL DAN PEMBAHASAN

Alur proses penelitian ini dibangun menggunakan RapidMiner Studio dengan pendekatan modular yang terstruktur dan dapat direplikasi. Gambar 1 menunjukkan tahapan yang dilakukan mulai dari pemuatan data mentah hingga evaluasi model. Proses diawali dengan Read CSV, yaitu pemanggilan dataset Telco Customer Churn yang memiliki 7.043 data pelanggan dan 21 atribut. Dataset ini kemudian dimasukkan ke modul Replace Missing Values untuk menangani nilai kosong. Berdasarkan eksplorasi awal, atribut TotalCharges memiliki beberapa nilai kosong yang diatasi dengan mengisi menggunakan nilai median. Pemilihan median dilakukan karena lebih tahan terhadap outlier dibandingkan rata-rata.



Gambar 1. Alur Proses di RapidMiner

Tahapan selanjutnya adalah mengatur peran (set role) menggunakan modul Set Role, di mana atribut Churn ditetapkan sebagai label (target variabel) untuk klasifikasi. Kemudian, karena distribusi kelas Churn menunjukkan ketidakseimbangan (hanya sekitar 26% pelanggan *churn*), dilakukan proses penyeimbangan data menggunakan SMOTE (Synthetic Minority Over-sampling Technique). Parameter yang digunakan adalah percentage = 200% dan  $k = 5$ , yang berarti data sintesis dibuat tiga kali lipat dari kelas minoritas untuk mendekati jumlah data pada kelas mayoritas. SMOTE ini sangat penting agar model tidak bias terhadap kelas *non-churn* dan mampu mengenali pola pelanggan *churn* secara lebih akurat.

Setelah penyeimbangan selesai, data dipisah menggunakan Split Data dengan proporsi 70% sebagai data latih dan 30% sebagai data uji. Proses pelatihan dilakukan dengan Logistic Regression, sedangkan hasilnya dievaluasi dengan modul Apply Model dan Performance. Evaluasi menggunakan metrik akurasi, precision, recall, dan AUC.

### 1. Hasil Evaluasi Model

Setelah model Logistic Regression dilatih dan diuji, diperoleh hasil yang sangat memuaskan. Berdasarkan confusion matrix dan metrik evaluasi yang ditampilkan dalam Gambar 2 dan Tabel 1–2, model menunjukkan performa klasifikasi yang sangat baik.

Tabel I  
 Confusion Matrix

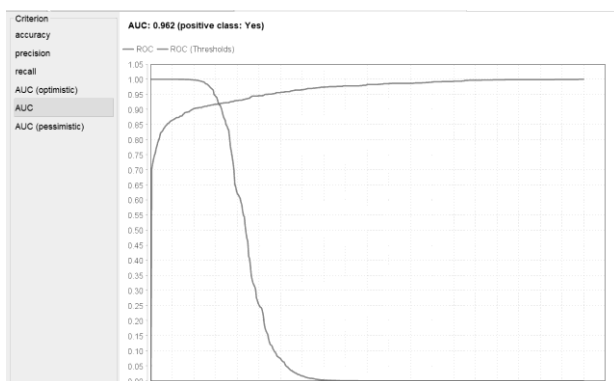
Prediksi / Aktual	No (Tidak Churn)	Yes (Churn)
No	1.482	218
Yes	70	1.334

Dari confusion matrix tersebut, dapat dilihat bahwa model mampu mengklasifikasikan 1.334 pelanggan churn dengan benar (true positive) dari total 1.552 pelanggan churn, dan mengklasifikasikan 1.482 pelanggan non-churn dengan benar (true negative) dari total 1.552 pelanggan non-churn. Kesalahan prediksi relatif kecil, dengan 218 false negative (churn tidak terdeteksi) dan hanya 70 false positive (bukan churn tapi diprediksi churn)

Tabel II  
 Metrik Evaluasi Model Logistic Regression

Metrik Evaluasi	Nilai
Akurasi	90.72%
Precision	95.01%
Recall	85.95%
AUC	0.962

Nilai akurasi sebesar 90.72% menunjukkan bahwa secara keseluruhan, model memiliki tingkat prediksi yang tinggi. Nilai precision sebesar 95.01% berarti bahwa dari seluruh prediksi churn yang dilakukan, sebanyak 95% benar-benar churn. Sedangkan recall sebesar 85.95% menunjukkan kemampuan model dalam menangkap pelanggan yang benar-benar akan berhenti berlangganan. Nilai AUC (Area Under Curve) sebesar 0.962 mengindikasikan bahwa model sangat baik dalam membedakan pelanggan churn dan non-churn (AUC mendekati 1 dianggap ideal).



Gambar 2. Hasil Evaluasi Akurasi, Precision, Recall dan Kurva ROC-AUC

## 2. Faktor Penyebab Churn

Dari model Logistic Regression yang dihasilkan, koefisien regresi menunjukkan beberapa variabel paling signifikan yang memengaruhi keputusan pelanggan untuk berhenti berlangganan layanan telekomunikasi. Secara

matematis, model Logistic Regression menggunakan formula sebagai berikut:

$$P(Y=1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

Dalam model ini,  $P(Y=1|X)$  mewakili probabilitas bahwa seorang pelanggan akan melakukan *churn* (berhenti menggunakan layanan atau produk Anda), mengingat serangkaian karakteristik mereka ( $X$ ). Nilai probabilitas ini akan selalu berada di antara 0 dan 1. Bagian pembilang dari rumus ini adalah angka 1, yang secara sederhana merupakan konstanta dalam perhitungan. Bagian penyebut,  $1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}$ , mungkin terlihat rumit, tetapi sebenarnya menggambarkan bagaimana berbagai faktor memengaruhi probabilitas *churn*. Mari kita lihat komponen-komponennya:

- $e$  adalah bilangan Euler, konstanta matematika penting yang merupakan dasar dari logaritma natural.
- $\beta_0$  adalah *intercept*, yang bisa diartikan sebagai probabilitas *churn* "dasar" ketika semua variabel independen lainnya bernilai nol.
- $\beta_1, \beta_2, \dots, \beta_n$  adalah koefisien regresi. Setiap koefisien ini mengukur seberapa besar dampak atau pengaruh dari variabel independen terkait ( $X_1, X_2, \dots, X_n$ ) terhadap probabilitas *churn*. Misalnya, jika  $\beta_1$  terkait dengan "usia pelanggan", nilai positif  $\beta_1$  berarti pelanggan yang lebih tua cenderung memiliki probabilitas *churn* yang lebih tinggi, sementara nilai negatif berarti probabilitasnya lebih rendah.
- $X_1, X_2, \dots, X_n$  adalah variabel independen atau fitur pelanggan yang Anda gunakan untuk memprediksi *churn*. Ini bisa berupa data demografi (usia, jenis kelamin), riwayat penggunaan (durasi langganan, frekuensi transaksi), perilaku (jumlah keluhan, interaksi dengan layanan pelanggan), atau faktor lain yang Anda yakini memengaruhi keputusan pelanggan untuk *churn*.

Secara keseluruhan, rumus ini mengambil kombinasi linear dari variabel-variabel independen pelanggan (setelah dikalikan dengan bobot atau koefisien masing-masing) dan mengubahnya melalui fungsi logistik (yang melibatkan  $e$ ) untuk menghasilkan probabilitas antara 0 dan 1. Semakin tinggi nilai dalam eksponen ( $-(\dots)$ ), semakin kecil penyebutnya, dan semakin besar probabilitas *churn*-nya. Sebaliknya, semakin rendah nilai dalam eksponen, semakin besar penyebutnya, dan semakin kecil probabilitas *churn*-nya. Dengan demikian, model ini membantu Anda memahami faktor-faktor mana yang paling berpengaruh dalam mendorong atau

mencegah pelanggan untuk *churn*, dan pada akhirnya memprediksi siapa yang berisiko tinggi untuk *churn*.

Dari hasil yang diperoleh, beberapa variabel utama yang paling berpengaruh antara lain:

1. Contract (Jenis Kontrak)

Pelanggan dengan kontrak month-to-month memiliki kemungkinan jauh lebih besar untuk *churn* dibandingkan mereka yang menggunakan kontrak tahunan. Kontrak jangka pendek menunjukkan fleksibilitas tinggi namun juga rendahnya loyalitas, sehingga menjadi indikator kuat *churn*.

2. Tenure (Lama Berlangganan)

Pelanggan yang baru berlangganan atau memiliki tenure pendek lebih rentan melakukan *churn*. Hal ini menunjukkan bahwa loyalitas pelanggan meningkat seiring dengan durasi berlangganan yang lebih lama. Oleh karena itu, fase awal berlangganan sangat krusial untuk menjaga kepuasan pelanggan.

3. MonthlyCharges (Biaya Bulanan)

Pelanggan dengan biaya bulanan tinggi memiliki kecenderungan lebih besar untuk *churn*, terutama bila biaya yang dibayar tidak sebanding dengan layanan yang diterima. Hal ini membuka peluang bagi perusahaan untuk melakukan pendekatan promosi atau diskon agar tetap kompetitif.

4. InternetService & TechSupport

Ketersediaan layanan internet dan dukungan teknis (tech support) juga memengaruhi *churn*. Pelanggan yang tidak mendapatkan layanan internet stabil atau tidak merasa terbantu oleh dukungan teknis cenderung beralih ke provider lain.

Dengan memahami variabel-variabel ini, perusahaan dapat menerapkan strategi mitigasi *churn* yang lebih efektif, misalnya dengan memberikan insentif untuk pelanggan dengan kontrak bulanan, meningkatkan kualitas layanan pelanggan baru, serta mengatur kembali struktur biaya bulanan berdasarkan perilaku pelanggan.

#### IV. KESIMPULAN

Penelitian ini berhasil membangun dan mengevaluasi model prediksi *churn* pelanggan pada perusahaan telekomunikasi menggunakan algoritma Logistic Regression di platform RapidMiner. Proses penelitian dilakukan secara sistematis dimulai dari tahap prapemrosesan data (preprocessing), penyeimbangan data menggunakan teknik SMOTE, hingga pelatihan dan evaluasi model dengan validasi silang. Dataset yang digunakan berasal dari Telco Customer Churn, yang

terdiri dari 7.043 entri data pelanggan dan 21 atribut, termasuk informasi demografis dan langganan layanan.

Hasil evaluasi model menunjukkan performa klasifikasi yang sangat baik. Model Logistic Regression mampu mencapai akurasi sebesar 90.72%, yang menunjukkan tingkat ketepatan prediksi model secara umum. Nilai precision sebesar 95.01% mengindikasikan bahwa sebagian besar prediksi *churn* yang dihasilkan benar-benar tepat. Sementara itu, recall sebesar 85.95% menunjukkan kemampuan model dalam mengenali pelanggan yang benar-benar berisiko *churn*. Nilai AUC sebesar 0.962 memperkuat validitas model, menandakan kemampuan diskriminatif yang tinggi antara pelanggan *churn* dan non-*churn*.

Dari analisis regresi, ditemukan bahwa variabel seperti jenis kontrak pelanggan (contract type), lama berlangganan (tenure), dan biaya bulanan (monthly charges) merupakan faktor yang paling berpengaruh terhadap perilaku *churn*. Pelanggan dengan kontrak bulanan, masa langganan yang pendek, serta tagihan bulanan yang tinggi memiliki kecenderungan lebih besar untuk berhenti berlangganan. Temuan ini sangat penting dalam menyusun strategi retensi yang lebih terarah dan berbasis data.

Secara keseluruhan, model ini dapat menjadi alat bantu yang efektif bagi perusahaan telekomunikasi dalam mengidentifikasi pelanggan berisiko *churn* secara proaktif. Dengan demikian, perusahaan dapat merancang intervensi dini seperti promosi, diskon, atau peningkatan layanan khusus kepada pelanggan yang berpotensi *churn*. Penerapan sistem prediksi *churn* seperti ini dapat membantu mengurangi kehilangan pelanggan, mempertahankan loyalitas pengguna, serta meningkatkan daya saing perusahaan di industri yang sangat kompetitif.

#### V. REFERENSI

- [1] H. Jain, A. Khunteta, and S. Srivastava, "Churn Prediction in Telecommunication using Logistic Regression and Logit Boost," *Procedia Computer Science*, vol. 167, pp. 101–112, 2020.
- [2] H. Nalattissifa and H. F. Pardede, "Customer Decision Prediction Using Deep Neural Network on Telco Customer Churn Data," *Jurnal Elektronika dan Telekomunikasi*, vol. 21, no. 2, pp. 122–127, 2021.
- [3] H. Donat and S. Karadayı Usta, "The Attitudes of the Telecommunication Customers in the COVID-19 Outbreak: The Effect of the Feature Selection Approach in Churn Analysis," *Sakarya University Journal of Science*, vol. 26, no. 3, pp. 530–544, 2022.
- [4] D. Kumari, S. K. Singh, S. S. Katira, I. V. Srinivas, and U. Salunkhe, "Telecom Customer Churn Forecasting Using Machine Learning: A Data-Driven Predictive Framework," *Metallurgical and Materials Engineering*, 2021.

- [5] B. Maulana, D. Febrian, I. R. Fachrezi, and M. F. Zeen, "Comparison of Support Vector Machine, Random Forest, and C4.5 Algorithms for Customer Loss Prediction," *Indonesian Journal of Applied Technology and Innovation Science*, vol. 2, no. 1, pp. 1–6, 2025.
- [6] S. K. Routray, "Marketing Strategy through Machine Learning Techniques: A Case Study at Telecom Industry," Telecom Training Center, Bhubaneswar, India, 2024.
- [7] S. Mitrović, B. Baesens, W. Lemahieu, and J. De Weerd, "On the operational efficiency of different feature types for telco Churn prediction," *European Journal of Operational Research*, vol. 267, no. 3, pp. 1141–1155, 2018.
- [8] D. H. Tisantri, R. C. Wihandika, and S. Adinugroho, "Prediksi Keputusan Pelanggan Menggunakan Extreme Learning Machine Pada Data Telco Customer Churn," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 11, pp. 10516–10523, 2019.
- [9] K. Kelvin, C. Cindy, C. Charles, D. P. Leonardo, and Y. Yennimar, "Customer Churn's Analysis In Telecommunications Company Using Fp-Growth Algorithm," *Jurnal Mantik*, vol. 4, no. 2, pp. 1285–1290, 2020.
- [10] R. Yu, X. An, B. Jin, J. Shi, O. A. Move, and Y. Liu, "Particle classification optimization-based BP network for telecommunication customer churn prediction," *Neural Computing and Applications*, vol. 29, pp. 707–720, 2018.
- [11] B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1414–1425, 2012.
- [12] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, "Applications of support vector machine (SVM) learning in cancer genomics," *Cancer Genomics and Proteomics*, vol. 15, no. 1, pp. 41–51, 2018.
- [13] H. Nalatissifa and H. F. Pardede, "Customer Decision Prediction Using Deep Neural Network on Telco Customer Churn Data," *Jurnal Elektronika dan Telekomunikasi*, vol. 21, no. 2, pp. 122–127, 2021.
- [14] N. Hashmi, N. A. Butt, and M. Iqbal, "Customer churn prediction in telecommunication a decade review and classification," *International Journal of Computer Science Issues*, vol. 10, no. 5, p. 271, 2013.
- [15] S. D. Damanik and M. I. Jambak, "Klasifikasi Customer Churn pada Telekomunikasi Industri Untuk Retensi Pelanggan Menggunakan Algoritma C4.5," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 1303–1309, 2023.