

**ANALISIS PREDIKSI CUSTOMER CHURN PADA LAYANAN
TELEKOMUNIKASI MENGGUNAKAN ALGORITMA
MACHINE LEARNING**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

NASYWA FEBIA HARTONO

22.11.5029

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS PREDIKSI CUSTOMER CHURN PADA LAYANAN
TELEKOMUNIKASI MENGGUNAKAN ALGORITMA
MACHINE LEARNING**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
NASYWA FEBIA HARTONO
22.11.5029

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS PREDIKSI CUSTOMER CHURN PADA LAYANAN TELEKOMUNIKASI MENGGUNAKAN ALGORITMA MACHINE LEARNING

yang disusun dan diajukan oleh

Nasywa Febia Hartono

22.11.5029

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal

Dosen Pembimbing,



Majid Rahardi, S.Kom, M.Eng
NIK. 190302393

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS PREDIKSI CUSTOMER CHURN PADA LAYANAN TELEKOMUNIKASI MENGGUNAKAN ALGORITMA MACHINE LEARNING

yang disusun dan diajukan oleh

Nasywa Febia Hartono

22.11.5029

Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Mei 2026

Susunan Dewan Penguji

Nama Penguji

Mulia Sulistiyono, S.kom.,
M.Kom.,
NIK. 190302248

Enda Putri Atika, M.Kom.
NIK. 190302706

Majid Rahardi, S.Kom.,
M.Eng.
NIK. 190302393

Tanda Tangan



Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Mei 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

Pang Luhendangtangan di Davran III,

NIM : 22.11.5029

Menyatakan bahwa Skripsi dengan judul berikut:

Dosen Pembimbing : Melid Rahanti, S.Kom, M.Eng.

Dosen Pembimbing : Majid Rahardi, S.Kom, M.Eng

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan **gagasan, rumusan dan penelitian SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 16 Juni 2026

Yang Menyatakan,



Nasywa Febia Hartono

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam kepada Allah SWT, penulis mempersembahkan karya sederhana ini kepada:

1. Orang tua tercinta, Bapak Eko Puji Hartono dan Ibu Henny Erawati yang selalu memberikan kasih sayang, doa, dukungan, pengorbanan, dan usaha tanpa henti untuk membantu penulis dalam menempuh pendidikan dan meraih kebahagiaan.
2. Kakak laki-laki tercinta, Fauzan Septiano Hartono yang selalu menjadi sumber motivasi dan telah mengajarkan saya arti kerja keras serta keberanian menghadapi kehidupan.
3. Sahabat-sahabat tercinta, Susan Sulastris, Childa, dan Dinda yang selalu mendukung saya dengan dorongan dan motivasi, serta mendampingi saya di setiap tahap penulisan skripsi ini hingga selesai.
4. Abdi yang senantiasa memberikan dukungan, semangat, perhatian, dan motivasi kepada penulis selama menjalani proses perkuliahan hingga penyusunan skripsi ini dapat terselesaikan dengan baik.
5. Seluruh pihak yang telah memberikan doa, bantuan, dan dukungan, baik secara langsung maupun tidak langsung.

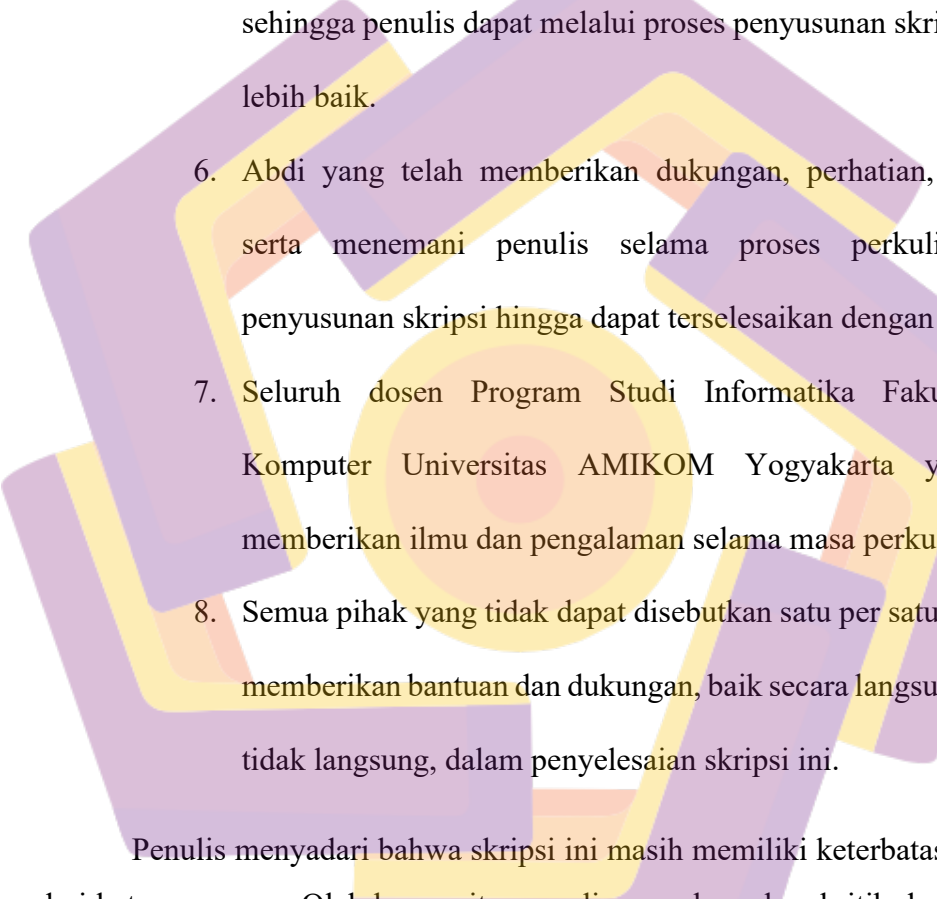
Semoga karya ini dapat menjadi langkah awal untuk terus belajar, berkembang, dan memberikan manfaat bagi banyak orang.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, serta hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Prediksi Customer Churn pada Layanan Telekomunikasi Menggunakan Algoritma Machine Learning” dengan baik.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta. Dalam proses penyusunannya, penulis memperoleh banyak bantuan, bimbingan, doa, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Allah SWT atas segala berkah, kesehatan, kemudahan, dan kekuatan yang diberikan selama seluruh proses penulisan skripsi ini.
2. Orang tua tercinta, Bapak Eko Puji Hartono dan Ibu Henny Erawati, atas doa, kasih sayang, pengorbanan, dukungan moral dan materi, serta dorongan yang tiada henti yang membantu saya menyelesaikan pendidikan ini.
3. Bapak Majid Rahardi, S.Kom., M.Eng. selaku dosen pembimbing yang telah memberikan arahan, ilmu, masukan, serta bimbingan dengan penuh kesabaran selama proses penyusunan skripsi.

- 
4. Kakak penulis, Fauzan Septiano Hartono, selalu memberikan motivasi dan dukungan agar penulis terus maju serta menghadapi tantangan dengan semangat.
 5. Sahabat-sahabat penulis, Susan Sulastri, Childa, dan Dinda, yang senantiasa memberikan dukungan, motivasi, serta bantuan sehingga penulis dapat melalui proses penyusunan skripsi dengan lebih baik.
 6. Abdi yang telah memberikan dukungan, perhatian, semangat, serta menemani penulis selama proses perkuliahan dan penyusunan skripsi hingga dapat terselesaikan dengan baik.
 7. Seluruh dosen Program Studi Informatika Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta yang telah memberikan ilmu dan pengalaman selama masa perkuliahan.
 8. Semua pihak yang tidak dapat disebutkan satu per satu yang telah memberikan bantuan dan dukungan, baik secara langsung maupun tidak langsung, dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan jauh dari kata sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca serta menjadi referensi bagi penelitian selanjutnya.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	i
HALAMAN PERSEMBAHAN	ii
KATA PENGANTAR	iii
DAFTAR ISI.....	v
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMBANG DAN SINGKATAN	xi
DAFTAR ISTILAH.....	xiii
INTISARI	xiv
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7

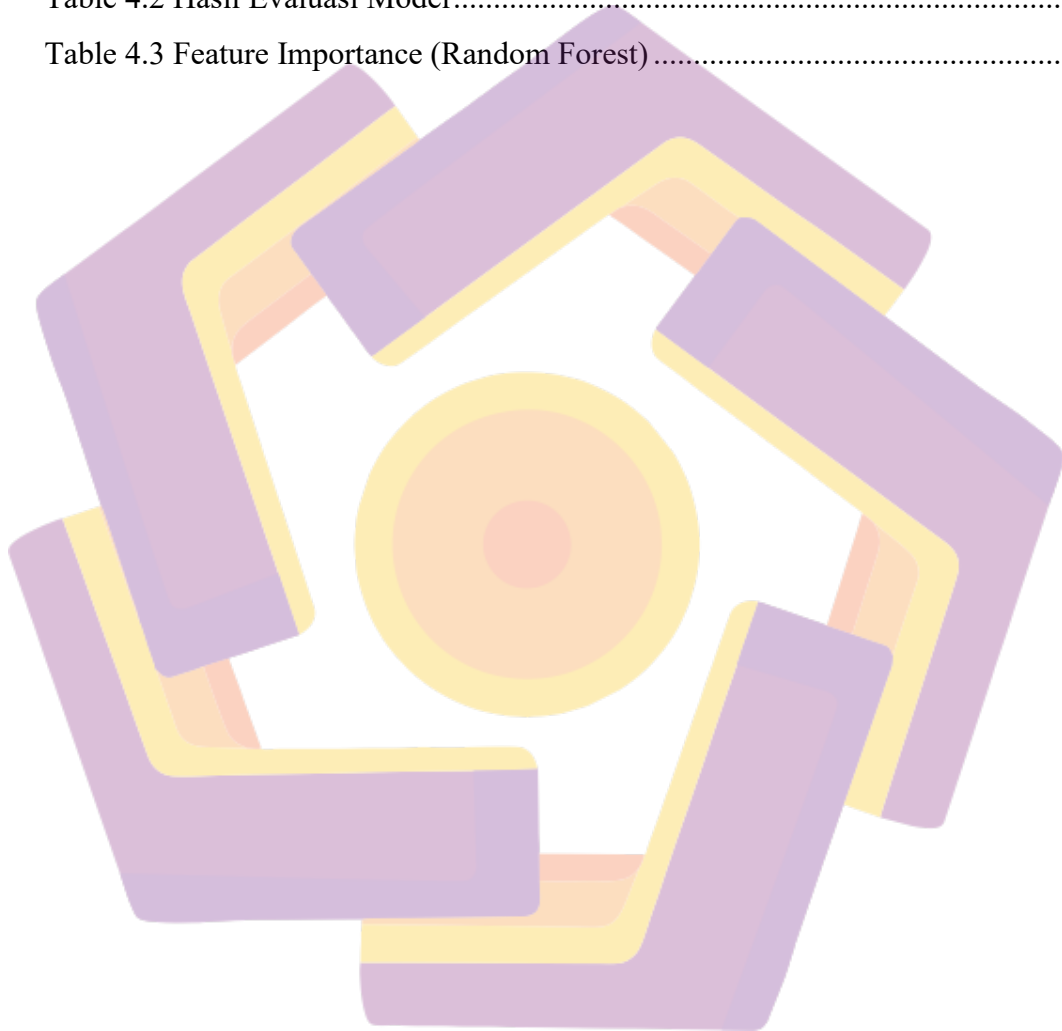
2.1	Studi Literatur	7
2.2	Dasar Teori.....	15
2.2.1	Customer Churn	15
2.2.2	Machine Learning	15
2.2.3	Klasifikasi	16
2.2.4	Logistic Regression.....	17
2.2.5	Random Forest	17
2.2.6	Gradient Boosting	18
2.2.7	Support Vector Machine (SVM).....	18
2.2.8	XGBoost	19
2.2.9	Confusion Matrix	20
2.2.10	Ukuran Evaluasi Model Klasifikasi	21
2.2.11	Interpretabilitas Model.....	22
BAB III METODE PENELITIAN		24
3.1	Objek Penelitian.....	24
3.2	Alur Penelitian	24
3.2.1	Akuisisi Dataset	25
3.2.2	Preprocessing	26
3.2.3	Split Data	27
3.2.4	Pelatihan Model	27
3.2.5	Evaluasi Model	27
3.2.6	Perbandingan Performa Model	27
3.2.7	Analisis dan Kesimpulan	28
3.2.8	Implementasi Prototipe Sistem	28

3.3 Alat dan Bahan.....	28
BAB IV HASIL DAN PEMBAHASAN	31
4.1 Akuisisi Dataset	31
4.2 Preprocessing Data.....	33
4.2.1 Penanganan Missing Value.....	33
4.2.2 Encoding Variabel Kategorikal.....	34
4.2.3 Feature Scaling	35
4.2.4 Analisis Distribusi Target	36
4.2.5 Exploratory Data Analysis (EDA).....	37
4.3 Split Data	47
4.4 Pelatihan Model	47
4.5 Evaluasi Model	50
4.5.1 Confusion Matrix sebagai Dasar Evaluasi.....	50
4.5.2 Accuracy	54
4.5.3 Precision.....	54
4.5.4 Recall	55
4.5.5 F1-Score.....	57
4.5.6 ROC-AUC	57
4.5.7 Calibration Curve XGBoost.....	59
4.5.8 Gain dan Lift Chart XGBoost.....	60
4.6 Perbandingan Performa Model	63
4.7 Interpretasi Model	64
4.7.1 Feature Importance (Random Forest)	65
4.7.2 Permutation Importance (Gradient Boosting).....	67

4.8 Analisis dan Kesimpulan	68
4.9 Implementasi Prototipe Sistem Prediksi Churn	69
4.9.1 Tujuan Pengembangan Prototipe	69
4.9.2 Arsitektur Sistem	70
4.9.3 Implementasi Model Terbaik	71
4.9.4 Tools yang Digunakan	71
4.9.5 Tampilan Sistem	72
4.9.6 Cara Kerja Sistem	76
4.9.7 Hasil Prediksi	76
4.9.8 Pengujian Sistem	77
4.9.9 Manfaat Prototipe dalam Konteks Bisnis	77
4.9.10 Deploy Online Menggunakan Streamlit	78
BAB V PENUTUP	79
5.1 Kesimpulan	79
5.2 Saran	80
REFERENSI	81
LAMPIRAN.....	83

DAFTAR TABEL

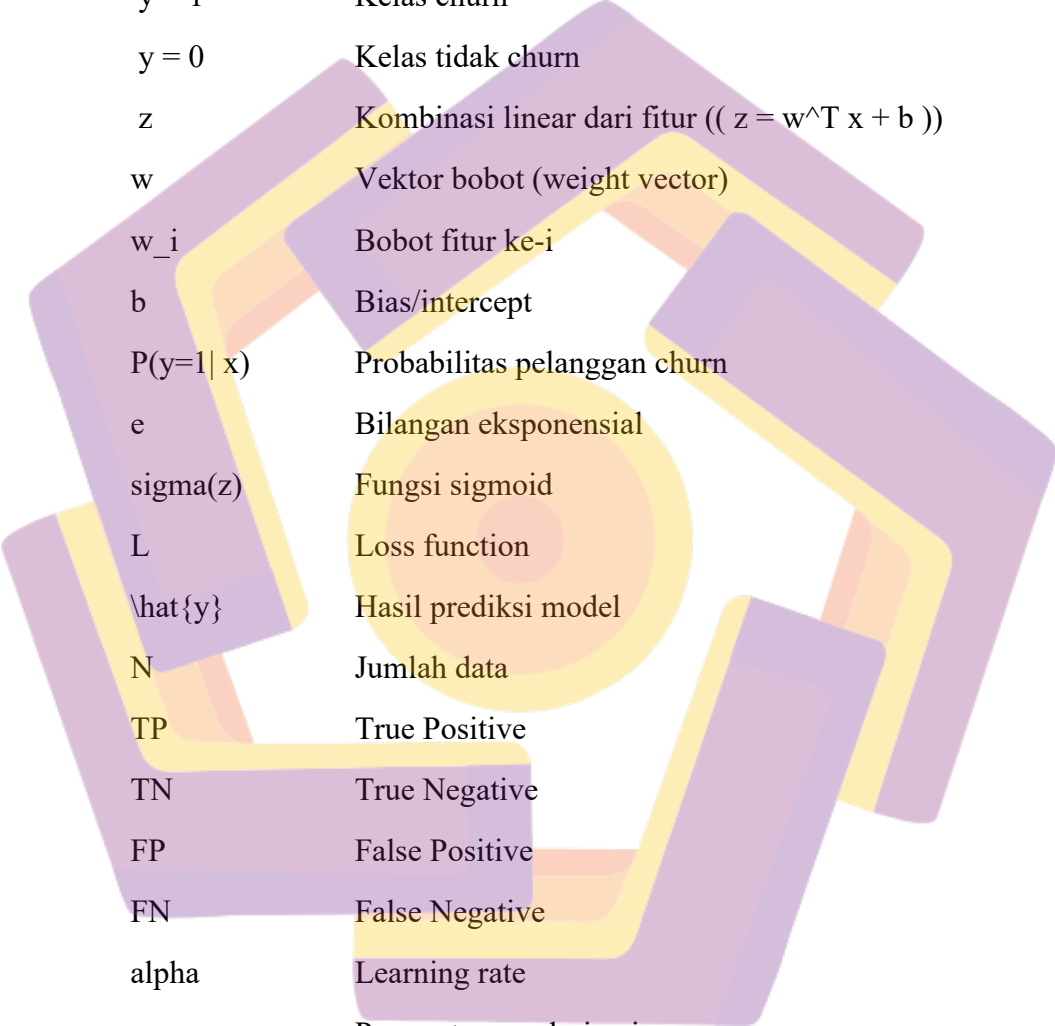
Table 2.1 Keaslian Penelitian	10
Table 2.2 Confusion Matrix	20
Table 4.2 Hasil Evaluasi Model.....	59
Table 4.3 Feature Importance (Random Forest)	65



DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	25
Gambar 4.1 Distribusi Churn	36
Gambar 4.2 Grafik Churn Berdasarkan Variabel Demografi	37
Gambar 4.3 Grafik Churn Rate Berdasarkan Jenis Kontrak	39
Gambar 4.4 Grafik Churn Rate Berdasarkan Jenis Internet Service	41
Gambar 4.5 Distribusi Tenure Pelanggan	42
Gambar 4.6 Distribusi Monthly Charges	42
Gambar 4.7 Boxplot Tenure Berdasarkan Status Churn	44
Gambar 4.8 Heatmap Korelasi Antar Fitur	45
Gambar 4.9 Confusion Matrix Gradient Boosting	52
Gambar 4.10 Confusion Matrix XGBoost	53
Gambar 4.11 Precision Recall Curve 5 Model	55
Gambar 4.12 ROC Curve Lima Model	58
Gambar 4.13 Calibration Curve XGBoost	60
Gambar 4.14 Gain Chart XGBoost	61
Gambar 4.15 Lift Chart XGBoost	62
Gambar 4.16 Grafik Perbandingan Accuracy Model	63
Gambar 4.17 Feature Importance (Random Forest)	66
Gambar 4.18 Permutation Importance (Gradient Boosting)	68
Gambar 4.19 Flow Sistem Prototipe Prediksi Customer Churn	70
Gambar 4.20 Tampilan Halaman Utama Sistem Prediksi Customer Churn	74
Gambar 4.21 Tampilan Hasil Prediksi Customer Churn	75

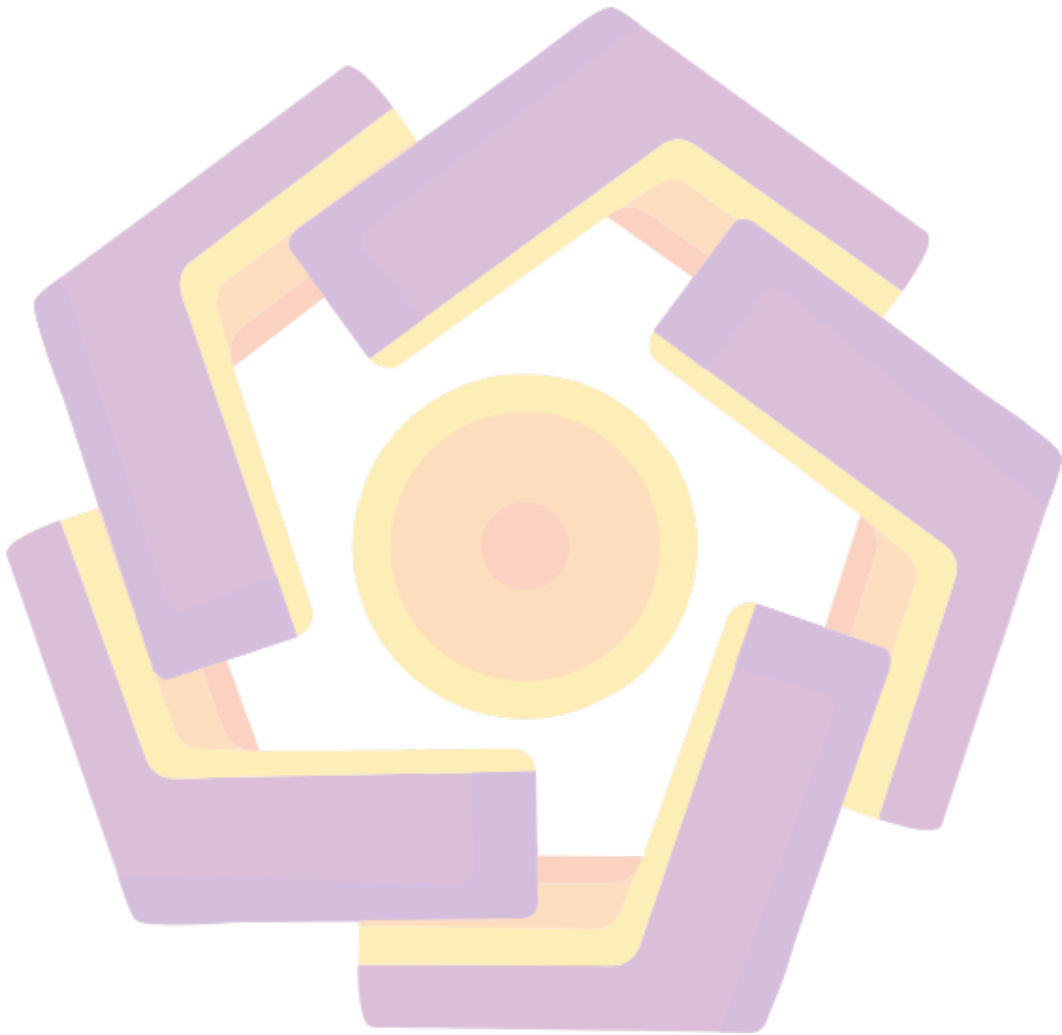
DAFTAR LAMBANG DAN SINGKATAN



x	Vektor fitur (feature vector)
x_i	Fitur ke-i
y	Label/kelas target
$y = 1$	Kelas churn
$y = 0$	Kelas tidak churn
z	Kombinasi linear dari fitur (($z = w^T x + b$))
w	Vektor bobot (weight vector)
w_i	Bobot fitur ke-i
b	Bias/intercept
$P(y=1 x)$	Probabilitas pelanggan churn
e	Bilangan eksponensial
$\sigma(z)$	Fungsi sigmoid
L	Loss function
$\hat{y}$	Hasil prediksi model
N	Jumlah data
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
α	Learning rate
γ	Parameter regularisasi
$f(x)$	Fungsi prediksi model
I_j	Importance fitur ke-j
Delta Accuracy	Penurunan akurasi pada Permutation Importance
ROC	Receiver Operating Characteristic

AUC

Area Under Curve



DAFTAR ISTILAH

Churn	Kondisi ketika pelanggan berhenti menggunakan layanan dan berpindah ke layanan lain.
Confusion Matrix	Tabel evaluasi yang digunakan untuk mengukur kinerja model klasifikasi berdasarkan hasil prediksi dan kondisi sebenarnya.
Customer Retention	Upaya perusahaan dalam mempertahankan pelanggan agar tetap menggunakan layanan yang diberikan.
Encoding	Proses mengubah data kategorikal menjadi data numerik agar dapat diproses oleh algoritma machine learning.
Exploratory Data Analysis (EDA)	Tahap analisis awal data untuk memahami karakteristik, pola, distribusi, dan hubungan antar variabel.
F1-Score	Metrik evaluasi yang merupakan rata-rata harmonis antara Precision dan Recall.
Feature	Variabel atau atribut yang digunakan sebagai masukan dalam proses pelatihan model machine learning.
Feature Importance	Metode interpretabilitas yang digunakan untuk mengukur tingkat pengaruh setiap fitur terhadap hasil prediksi model.
Hyperparameter	Parameter yang ditentukan sebelum proses pelatihan model dan digunakan untuk mengontrol proses pembelajaran algoritma.
Permutation Importance	Metode interpretabilitas yang mengukur pengaruh fitur dengan mengacak nilai fitur dan mengamati perubahan performa model.
StandardScaler	Metode scaling yang mengubah data sehingga memiliki rata-rata 0 dan standar deviasi 1.

INTISARI

Industri telekomunikasi terus mengamati tingkat churn pelanggan karena pelanggan berdampak pada penurunan pendapatan dan efektivitas upaya retensi pelanggan. Dengan menggunakan algoritma machine learning berbasis data tabular, penelitian ini bertujuan untuk membangun model yang dapat memprediksi tingkat penurunan pelanggan.

Dataset pelanggan telekomunikasi mencakup fitur dan pola pembayaran seperti tenure, TotalCharges, MonthlyCharges, jenis kontrak, layanan internet dan metode pembayaran, dengan pembagian data 80:20. Algoritma yang digunakan adalah Logistic Regression, Random Forest, Support Vector Machine (SVM), Gradient Boosting dan XGBoost. Analisis model didukung oleh analisis calibration curve, gain dan lift chart, serta ROC-AUC, accuracy, precision, dan recall. Features importance dan permutation importance digunakan untuk penyesuaian threshold dan analisis interpretabilitas.

Hasil penelitian menunjukkan bahwa algoritma XGBoost menunjukkan kinerja terbaik dibandingkan algoritma lainnya dengan akurasi sebesar 80,27% dan ROC-AUC sebesar 0,8415. Gradient Boosting menempati posisi kedua dengan akurasi sebesar 80,13% dan ROC-AUC sebesar 0,8411. Random Forest dan SVM memiliki kinerja yang lebih rendah dibandingkan model Logistic Regression, yang memiliki recall tertinggi sebesar 78,07% dan ROC-AUC sebesar 0,8412. Hasil ini menunjukkan algoritma berbasis boosting lebih baik menemukan pola kompleks dalam data churn.

Selain mengidentifikasi pelanggan yang berisiko berhenti berlangganan, teknik prediksi churn yang dihasilkan ini juga memberikan informasi mengenai faktor-faktor yang memengaruhi churn. Untuk meningkatkan retensi pelanggan di sektor telekomunikasi, model ini dapat digunakan sebagai landasan dalam pengambilan keputusan strategis.

Kata kunci: churn pelanggan, machine learning, interpretabilitas model, threshold, retensi pelanggan

ABSTRACT

The telecommunications industry continues to monitor customer churn rates, as churn impacts revenue and the effectiveness of customer retention efforts. Using machine learning algorithms based on tabular data, this study aims to build a model capable of predicting customer churn rates.

The telecommunications customer dataset includes features and payment patterns such as tenure, TotalCharges, MonthlyCharges, contract type, internet services, and payment methods, with an 80:20 data split. The algorithms used are Logistic Regression, Random Forest, Support Vector Machine (SVM), Gradient Boosting, and XGBoost. Model analysis is supported by calibration curve analysis, gain and lift charts, as well as ROC-AUC, accuracy, precision, and recall. Feature importance and permutation importance are used for threshold adjustment and interpretability analysis.

The results of the study show that the XGBoost algorithm performed best compared to the other algorithms, with an accuracy of 80.27% and an ROC-AUC of 0.8415. Gradient Boosting ranked second with an accuracy of 80.13% and an ROC-AUC of 0.8411. Random Forest and SVM performed worse than the Logistic Regression model, which had the highest recall at 78.07% and an ROC-AUC of 0.8412. These results indicate that boosting-based algorithms are better at identifying complex patterns in churn data.

In addition to identifying customers at risk of churning, this churn prediction technique also provides insights into the factors that influence churn. To improve customer retention in the telecommunications sector, this model can serve as a foundation for strategic decision makings.

Keyword: customer churn, machine learning, model interpretability, threshold, customer retention.