

Analisis dan Prediksi *Customer Churn* pada *Platform Streaming* Berbasis Langganan Menggunakan Metode *Random Forest*

**Imakulata Kresnawati M Bili¹, I Wayan Sudiarta^{2*}, Maria Yuditia
Wungabelen³, Ni Kadek Alika Rosdiana⁴, Putri Rafiana⁵**

¹⁻⁵Program Studi Bisnis Digital, Institut Bisnis dan Teknologi Indonesia, Indonesia

**Penulis Korespondensi:* sudiarsa@instiki.ac.id

Abstract. *Customer churn is a strategic challenge for digital streaming platforms because it directly Impacts revenue and business sustainability. This study aims to analyze the factors influencing customer Churn and develop a churn prediction model using the Random Forest algorithm. The study uses a Quantitative approach with an explanatory design and utilizes secondary data from the Netflix Customer Churn and Engagement Dataset available on Kaggle. The dataset consists of 1,000 customer data with 16 Variables covering demographic characteristics, service usage behavior, financial condition, and customer Satisfaction level. The data was processed through preprocessing, one-hot encoding, and a 70:30 split Between training and test data. Model performance was evaluated using accuracy, precision, recall, F1 Score, and ROC-AUC metrics. The results show that the Random Forest model produces an accuracy of 53.7%, precision of 56.3%, recall of 63.6%, F1-score of 59.7%, and ROC-AUC of 0.534, indicating Moderate predictive ability and only slightly better than random classification. Feature importanceAn.evealed that user engagement levels, such as viewing duration and frequency of interactions, Were the most dominant factors influencing churn, followed by economic factors and customer satisfaction. The results of this study are expected to provide a basis for streaming platforms to design more effective Customer retention strategies.*

Keywords: *Customer Churn; Customer Retention; Machine Learning; Random Forest; Streaming Platform.*

Abstrak. Customer churn merupakan tantangan strategis bagi platform streaming digital karena, berdampak langsung terhadap pendapatan dan keberlanjutan bisnis. Penelitian ini bertujuan untuk Menganalisis faktor-faktor yang memengaruhi churn pelanggan serta mengembangkan model Prediksi churn menggunakan algoritma Random Forest. Penelitian menggunakan pendekatan Kuantitatif dengan desain eksplanatif dan memanfaatkan data sekunder dari Netflix Customer Churn and Engagement Dataset yang tersedia di Kaggle. Dataset terdiri dari 1.000 data pelanggan Dengan 16 variabel yang mencakup karakteristik demografis, perilaku penggunaan layanan. Kondisi keuangan, dan tingkat kepuasan pelanggan. Data diproses melalui tahapan preprocessing, One-hot encoding, serta pembagian data latih dan data uji dengan rasio 70:30. Kinerja model Dievaluasi menggunakan metrik akurasi, precision, recall, F1-score, dan ROC-AUC. Hasil Penelitian menunjukkan bahwa model Random Forest menghasilkan akurasi sebesar 53,7%, Precision 56,3%, recall 63,6%, F1-score 59,7%, dan ROC-AUC 0,534, yang menunjukkan kemampuan prediksi pada tingkat sedang dan hanya sedikit lebih baik dibandingkan klasifikasi Acak. Analisis feature importance mengungkapkan bahwa tingkat keterlibatan pengguna, seperti Durasi menonton dan frekuensi interaksi, merupakan faktor paling dominan dalam memengaruhi Churn, diikuti oleh faktor ekonomi dan kepuasan pelanggan. Hasil penelitian ini diharapkan dapat Menjadi dasar bagi platform streaming dalam merancang strategi retensi pelanggan yang lebih Efektif.

Kata kunci: *Churn Pelanggan; Machine Learning; Platform Streaming; Random Forest; Retensi Pelanggan.*

1. LATAR BELAKANG

Churn adalah fenomena yang terjadi ketika pelanggan berhenti menggunakan atau membicarakan suatu layanan setelah jangka waktu tertentu. Dalam dunia bisnis, churn berfungsi sebagai indikator penting pendapatan dan pertumbuhan perusahaan. Penelitian menunjukkan bahwa meningkatnya tingkat churn pelanggan berhubungan dengan tingkat kepuasan pelanggan yang lebih tinggi, tingkat ketidakpuasan pengguna yang lebih rendah, dan tingkat ketidakpuasan pengguna yang lebih rendah terhadap layanan yang diberikan serta biaya

yang dibebankan kepada pengguna (Ullah et al., 2019) (Ahmad et al., 2019; (Beeharry & Fokone, 2021;). Oleh karena itu, mengidentifikasi pelanggan yang berpotensi berhenti berlangganan telah menjadi persyaratan strategis bagi layanan digital.

Streaming digital mengumpulkan data pengguna yang memengaruhi penggunaan layanan. Data menunjukkan durasi akses, frekuensi penggunaan, tingkat konsumsi, dan interaksi dengan penyedia layanan (Chang et al., 2024; Zhou et al., 2023). Penelitian telah menunjukkan bahwa keterlibatan pengguna memiliki hubungan yang signifikan dengan churn, dengan peningkatan aktivitas dan kepuasan pengguna sebagai indikator utama retensi pengguna. Misalnya, karakteristik demografis dan keuangan pelanggan memengaruhi persepsi mereka tentang harga dan tingkat layanan (Muneer et al., 2022 ; Thakkar et al., 2022).

2. KAJIAN TEORITIS

Mengocok Pengurangan jumlah pelanggan (customer) adalah suatu fenomena yang terjadi yang terjadi ketika pelanggan berhenti menggunakan atau membicarakan suatu layanan setelah jangka waktu tertentu. Dalam dunia bisnis, tingkat churn merupakan indikator penting bagi pendapatan dan pertumbuhan perusahaan. Penelitian telah menunjukkan sebelumnya itu bahwa meningkatnya tingkat churn pelanggan meningkat dengan meningkatnya kepuasan pelanggan, loyalitas pengguna, dan keseimbangan antara layanan yang diberikan dan biaya yang dibebankan kepada pelanggan. Tingkat pergantian pelanggan (*customer churn*) terkait dengan peningkatan kepuasan pelanggan, loyalitas pengguna, dan keseimbangan antara layanan yang diberikan dan biaya yang dibebankan kepada pelanggan. Oleh dari hal ini, mengidentifikasi pelanggan yang akan berhenti berlangganan. Potensi telah menjadi persyaratan strategis bagi penyedia layanan digital.

Platform streaming mengumpulkan data pengguna yang memengaruhi penggunaan layanan. Mengumpulkan data pengguna yang memengaruhi penggunaan layanan. Data ini menunjukkan durasi akses, frekuensi penggunaan, tingkat konsumsi, dan interaksi dengan penyedia layanan (Chang et al., 2024 ; Zhou et al., 2023). Penelitian telah menunjukkan sebelumnya itu bahwa keterlibatan pengguna memiliki hubungan yang signifikan dengan tingkat churn, dengan peningkatan aktivitas dan kepuasan pengguna sebagai indikator utama retensi pengguna. Keterlibatan pengguna memiliki hubungan yang signifikan dengan tingkat churn. Sebagai contoh, demografi karakteristik dan karakteristik keuangan pelanggan memengaruhi persepsi mereka terhadap harga dan tingkat layanan (Muneer et al., 2022; Thakkar et al ., 2022).

3. METODE PENELITIAN

Data Penelitian dan Karakteristik Data

Studi ini menggunakan pendekatan kuantitatif dengan desain penelitian eksplanatif untuk menganalisis faktor-faktor yang mempengaruhi pengurangan pelanggan dan mengembangkan model prediksi churn menggunakan algoritma Random Forest. Data yang digunakan adalah data sekunder yang diambil dari Netflix Customer Churn and Engagement Dataset yang tersedia di Kaggle. Kumpulan data terdiri dari 1.000 titik data pelanggantik data dengan 16 variabel yang mengukur karakteristik demografis, penggunaan layanan, status keuangan, dan indikator kepuasan pelanggan (Ullah et al., 2019). 16 variabel yang mengukur karakteristik demografis, penggunaan layanan, status keuangan, dan indikator kepuasan pelanggan (Ullah et al., 2019). Variabel dependen dalam penelitian ini adalah status churn, yang menunjukkan apakah peserta terlibat atau tidak. Seperti Xie et al. (2009), variabel independen meliputi jumlah waktu yang dihabiskan di harian, tingkat loyalitas pengguna, pendapatan bulanan, usia, skor kepuasan pelanggan, jumlah permintaan dukungan, jumlah profil, lama berlangganan, dan riwayat pembayaran yang dibayarkan (Xie et al., 2009). Teknik pengumpulan data dilakukan melalui dokumentasi dataset yang tersedia untuk umum. Pada tahap pengumpulan data, dilakukan perbandingan antara variabel yang sesuai dan variabel target. Variabel kategorikal dienkode menggunakan metode one-hot encoding sehingga dapat dibaca oleh model pembelajaran mesin. Setelah itu, data dibagi menjadi dua kategori: data latihan dan data uji, dengan 70% untuk data latihan dan 30% untuk data uji (Shu-li et al., 2021). Karakteristik ini memungkinkan pemahaman yang lebih mendalam tentang bagaimana faktor-faktor yang disebutkan di atas memengaruhi kesediaan pelanggan untuk tetap bertahan dalam jangka panjang, yang lebih mendalam tentang bagaimana berbagai faktor berkontribusi pada kemampuan pengguna untuk memahami situasi tersebut.

Metode Pengumpulan dan Pengolahan Data

Teknik pengumpulan data dilakukan melalui dokumentasi dataset yang tersedia untuk umum. Selama fase pengumpulan data, dilakukan perbandingan antara variabel target dan fitur. One-hot encoding digunakan untuk mentransformasi variabel kategorikal sehingga model pembelajaran mesin dapat menganalisisnya. Data dipisahkan menjadi data latih dan uji, dengan rasio data latih 70% dan rasio data uji 30% (Shu-li et al., 2021). Tahapan pra-pemrosesan juga mendorong pengumpulan data untuk mengidentifikasi Entri yang singkat atau tidak relevan. Proses ini bertujuan untuk memastikan bahwa data yang digunakan dalam model berkualitas tinggi dan sesuai untuk analisis yang lebih mendalam (Gao et al., 2019).

```
import pandas as pd

df = pd.read_csv('/content/netflix_large_user_data.csv')
print("First 5 rows of the dataset:")
print(df.head())

First 5 rows of the dataset:
   Customer ID  Subscription Length (Months) \
0          C00001                           12
1          C00002                           12
2          C00003                            3
3          C00004                            3
4          C00005                          24

   Customer Satisfaction Score (1-10)  Daily Watch Time (Hours) \
0                           10                  4.85
1                            8                  1.75
2                            4                  2.75
3                            7                  3.00
4                            2                  1.37

   Engagement Rate (1-10)  Device Used Most Often Genre Preference \
0                           4                  Tablet        Action
1                           9                  Laptop      Thriller
2                           9                 Smart TV     Comedy
3                           9                 Smart TV     Drama
4                           5                  Mobile     Drama

   Region Payment History (On-Time/Delayed)  Subscription Plan \
0        Europe          On-Time            Basic
1        Europe          On-Time            Basic
2        Asia            On-Time            Premium
3        Europe          Delayed            Premium
4  North America          On-Time            Standard

   Churn Status (Yes/No)  Support Queries Logged  Age  Monthly Income ($)
0            No                  10    33          6250
1            Yes                  9    28          7018
2            Yes                  3    18          1055
3            No                  5    32          6707
4            Yes                  2    59          1506

   Promotional Offers Used  Number of Profiles Created
0                           5                           2
1                           1                           5
2                           1                           5
3                           5                           4
4                           3                           5
```

Gambar 1. Kumpulan Data Dan Pemahaman Data Awal (Sumber Google Colab).

Model dan Alat Analisis Data

Model prediksi churn yang digunakan dalam penelitian ini adalah Random Forest Classifier, sebuah teknik pembelajaran ensemble berbasis pohon keputusan. Random Forest dipilih karena kemampuannya dalam menangani data berdimensi tinggi, mengurangi risiko overfitting, serta menyediakan informasi tingkat kepentingan variabel (feature important) untuk membantu dalam interpretasi hasil analisis (Zhang et al., 2021; Chang et al., 2024). Analisis dilakukan menggunakan perangkat lunak pengolahan data berbasis Python seperti Pandas untuk manipulasi data, Scikit-learn untuk pemodelan pembelajaran mesin, dan Matplotlib serta Seaborn untuk visualisasi data. Model tersebut diuji menggunakan data latihan dan kemudian

dievaluasi menggunakan data uji untuk menilai kinerja model dalam memprediksi pelanggan yang akan berhenti berlangganan (Öskarsdóttir dkk., 2017). Tujuan dari penggunaan berbagai pendekatan ini adalah untuk mendapatkan pemahaman komprehensif tentang kemampuan model dalam mengidentifikasi pelanggan yang berisiko berhenti berlangganan (Edwine et al., 2022; Văduva et al., 2024).

```

X = df.drop(columns=['Customer ID', 'Churn Status (Yes/No)'])
y = df['Churn Status (Yes/No)']

# Check for missing values in X
missing_values_X = X.isnull().sum()
missing_values_X = missing_values_X[missing_values_X > 0]

# Check for missing values in y
missing_values_y = y.isnull().sum()

print("\nMissing values in features (X):")
if missing_values_X.empty:
    print("No missing values found in X.")
else:
    print(missing_values_X.to_string())

print("\nMissing values in target (y):")
if missing_values_y == 0:
    print("No missing values found in y.")
else:
    print(f"Missing values in y: {missing_values_y}")

```

Missing values in features (X):
No missing values found in X.

Missing values in target (y):
No missing values found in y.

Gambar 2. Pemrosesan Data (Sumber Google Colab).

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, stratify=y, random_state=42)

print(f"Shape of X_train: {X_train.shape}")
print(f"Shape of X_test: {X_test.shape}")
print(f"Shape of y_train: {y_train.shape}")
print(f"Shape of y_test: {y_test.shape}")

Shape of X_train: (700, 26)
Shape of X_test: (300, 26)
Shape of y_train: (700,)
Shape of y_test: (300,)
```

Gambar 3. Pemisahan Data Untuk Pemodelan (Sumber Google Colab).

```
from sklearn.ensemble import RandomForestClassifier

# Initialize the Random Forest Classifier with a fixed random_state
rf_model = RandomForestClassifier(random_state=42)

# Train the model using the training data
rf_model.fit(X_train, y_train)

print("Random Forest Classifier model trained successfully.")

Random Forest Classifier model trained successfully.
```

Gambar 4. Pembuatan Dan Pelatihan Model Random Forest.

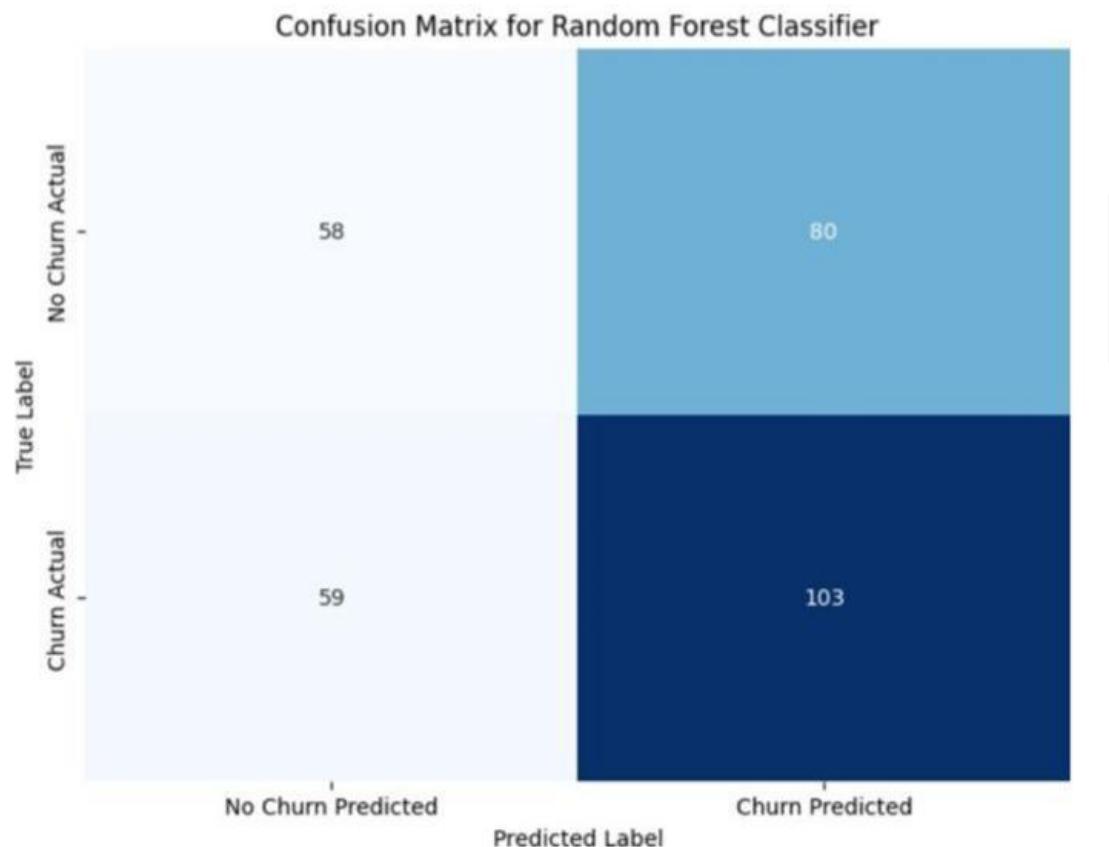
Evaluasi Kinerja Model

Beberapa metrik klasifikasi, termasuk akurasi, tekanan, recall, F1-score, dan ROC -AUC, digunakan untuk mengevaluasi kinerja model. Metrik akurasi memberikan informasi umum tentang beberapa prediksi penting, sementara metrik tekanan dan recall memberikan informasi yang lebih detail tentang kinerja model dalam mengidentifikasi pelanggan secara terarah. Dengan menganalisis hasil evaluasi model, kita dapat menentukan efektivitas algoritma Random Forest dalam konteks prediksi churn. Selain itu, hasil ini berfungsi sebagai dasar untuk analisis implikasi penelitian terhadap strategi retensi pelanggan pada platform streaming. Dengan menggunakan wawasan, penyedia layanan dapat menentukan tindakan dan intervensi terbaik untuk meningkatkan kepuasan pelanggan dan mengurangi tingkat kehilangan pelanggan (Arshad et al., 2022) dari wawasan, penyedia layanan dapat menentukan tindakan dan intervensi terbaik untuk meningkatkan kepuasan pelanggan dan mengurangi tingkat kehilangan pelanggan (Arshad et al., 2022).

4. HASIL DAN PEMBAHASAN

Evaluasi Kinerja Model

Prediksi dibuat berdasarkan data masukan dan dievaluasi menggunakan matriks kebingungan dan metrik klasifikasi .dibuat berdasarkan data masukan dan dievaluasi menggunakan matriks kebingungan,metrik klasifikasi



Gambar 5. Matrix Pengklasifikasian.

Berdasarkan matriks kebingungan, hasilnya adalah sebagai berikut. Berdasarkan padamatriks kebingungan, hasilnya disajikan matriks kebingungan. Hasilnya disajikan di bawah ini:

- Ada 103 BenarPositives (TNP). Hal positif.
- True Negative (TN) Negatif (TN): 58.
- Terdapat 80 False terdapat 80 hasil positif palsu (FP).).
- 59 False negatif (FN).

Matriks kebingungan menunjukkan bahwa menunjukkan bahwa model memiliki jumlah false positives dan false negatives yang tinggi, yang berdampak pada efektivitas strategi retensi pelanggan. memiliki jumlah false positive dan false negative yang tinggi, yang berdampak pada efektivitas strategi retensi pelanggan.

Metrik untuk evaluasi Model

Metrik	Nilai
Accuracy	0,537
Precision	0,563
Recall	0,636
F1-Score	0,597
ROC-AUC	0,534

Gambar 6. Tabel Hasil Evaluasi Model Random Forest.

Tabel Metrik Evaluasi Model Evaluasi Model Random Forest Model Random Forest Classifier yang dikembangkan untuk memprediksi pergantian pelanggan menghasilkan karakteristik kinerja berikut berdasarkan data :

a) *Akurasi (0,537)*

Model ini dapat memprediksi tingkat kehilangan pelanggan dengan akurasi sekitar 53,7 %. Angka tersebut sedikit lebih tinggi daripada tebakan acak, yang memiliki probabilitas 0,50 ketika mengkategorikan suatu masalah dengan jumlah variabel yang tetap. Distribusi nilai yang relatif merata (53,9% pelanggan berhenti berlangganan dibandingkan 46,1% pelanggan tetap), tingkat akurasi 50% menunjukkan bahwa prediksi model semakin memburuk.

b) *Presisi (0,563)*

Ketika Model tersebut memprediksi bahwa pelanggan akan berhenti berlangganan, dan prediksi tersebut memiliki akurasi 56,3%. bahwa ada sebagian besar pelanggan yang diprediksi akan berhenti berlangganan tetapi sebenarnya tidak ada. Dari sudut pandang bisnis, kondisi ini dapat menghasilkan peningkatan arus kas sebagai akibat dari insentif retensi yang diberikan kepada pelanggan yang cenderung tidak akan beralih ke pesaing.

c) *Recall (0,636)*

Model tersebut berhasil mengidentifikasi 63,6% dari seluruh pengguna yang kemungkinan akan mengalami churn. kata lain, model tersebut gagal mendeteksi sekitar 36,4% pelanggan yang berhenti berlangganan. Dari sudut pandang bisnis, kegagalan untuk mengidentifikasi pelanggan yang berpotensi berhenti berlangganan. Pelanggan yang berhenti dapat mengakibatkan kurangnya waktu untuk melakukan intervensi guna mempertahankan mereka agar tidak berhenti berlangganan.

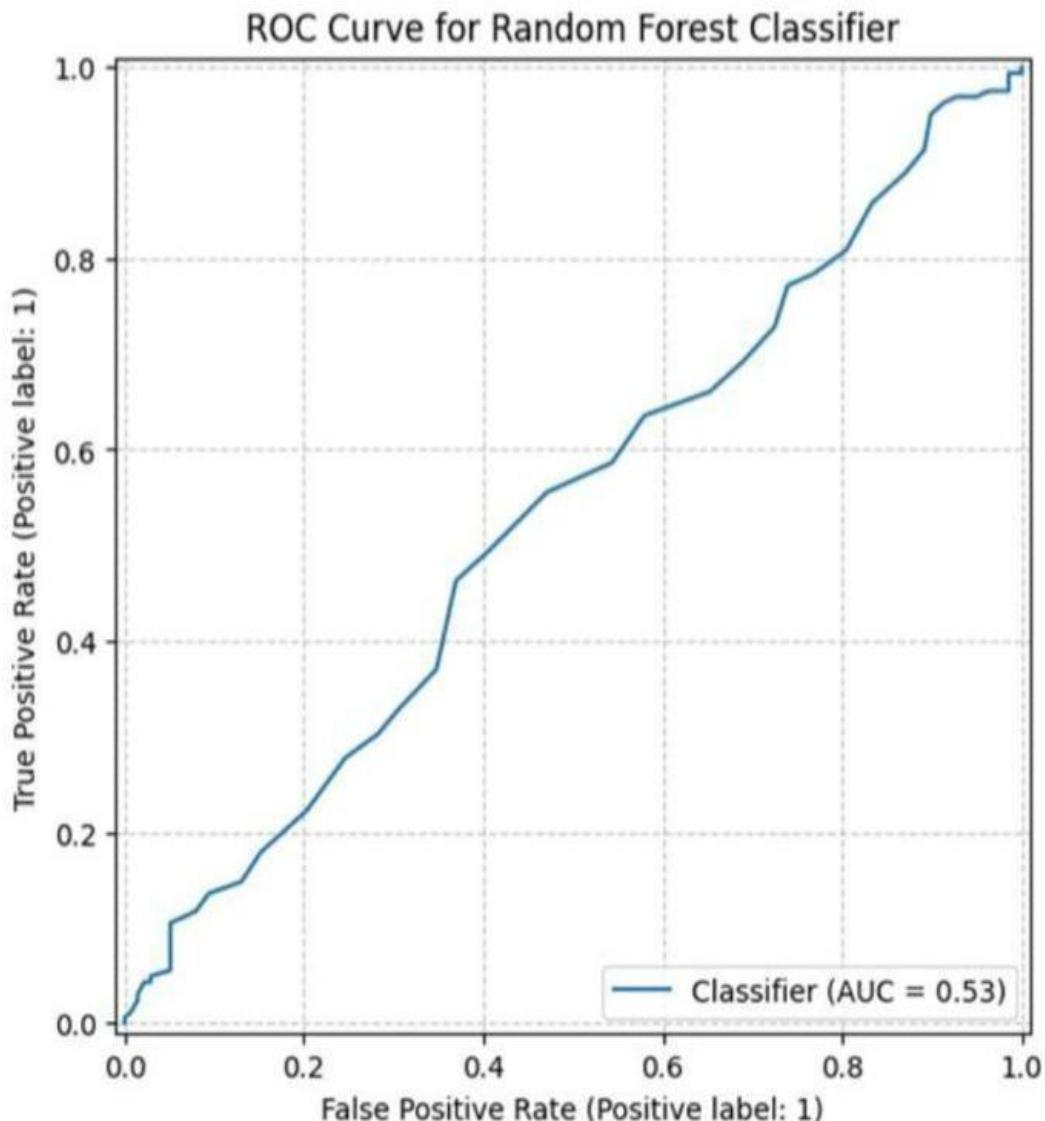
d) *Skor F1 skor(0,597)(0,597)*

Skor F1 adalah keseimbangan antara di antaradan daya ingat .tekanan dan ingatan. Skor sebesar 0,597 menunjukkan efektivitas model yang sedang mengklasifikasikan kelas positif (churn), namun tidak dapat dikategorikan sebagai model dengan prediksi kinerja yang kuat, diklasifikasikan sebagai kelas positif (churn), tidak dapat dikategorikan sebagai model dengan kinerja prediksi yang kuat.

e) *ROC-AUC(0.534)*

Area Nilai Area Under the Receiver Operating Characteristic Curve (ROC-AUC) sebesar 0,534 menunjukkan kemampuan model untuk membedakan antara pelanggan yang berhenti berlangganan (churn) dan pelanggan yang berhenti berlangganan (churn) kemampuan model Dan membedakan antara pelanggan yang akan berhenti berlangganan (churn) dan yang tidak pada berbagai klasifikasi.tidak mengalami pergantian pelanggan pada berbagai klasifikasi. ROC Nilai ROC -AUC sebesar 0,5 menunjukkan pekerjaan normal, sedangkan nilai 1 menunjukkan klasifikasi yang ketat. Nilai 0,5 menunjukkan pekerjaan normal, sedangkan nilai 1 menunjukkan klasifikasi yang ketat,yang unggul.Dengan nilai nilai yang jauh lebih rendah dari 0,5 , yaitu ini memiliki tingkat diskriminasi yang tinggi .Dengan nilai yang jauh lebih rendah dari 0,5, model ini memiliki tingkat diskriminasi yang tinggi.

Analisis Kurva ROC Kurva Melengkung



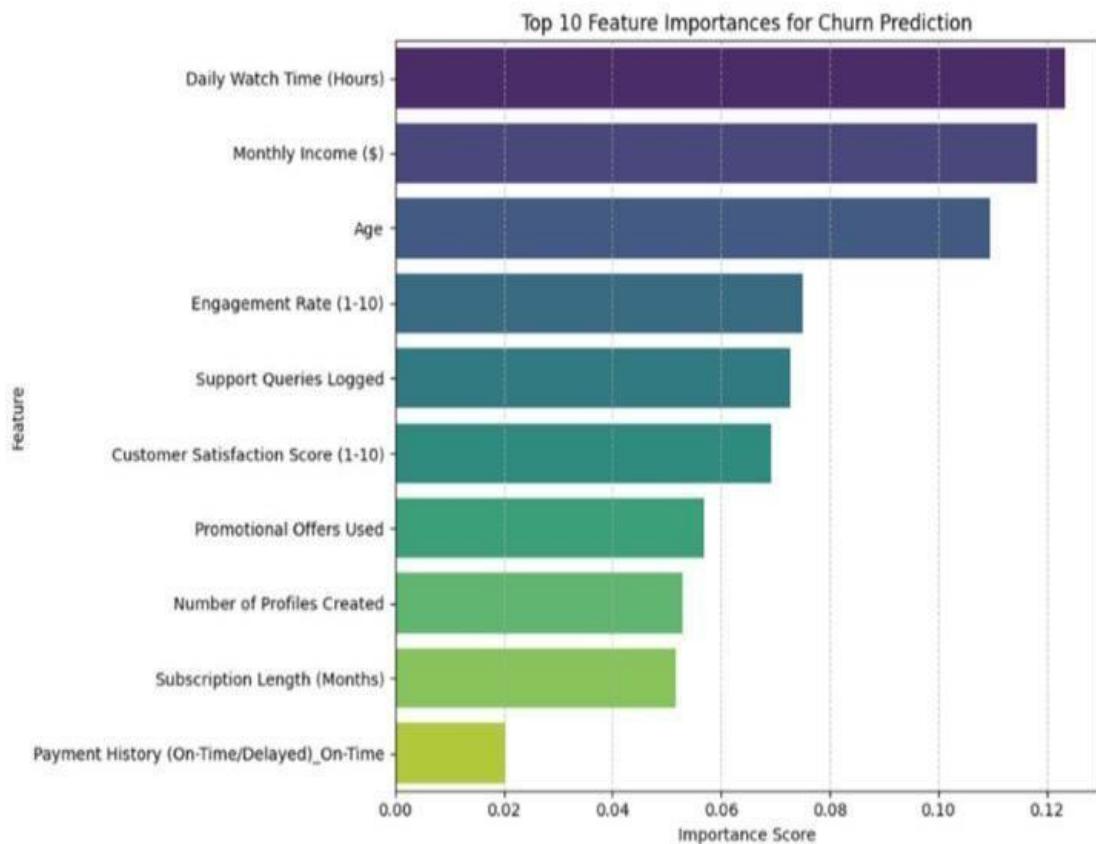
Gambar 7. Kurva ROC Klasifikasi Biner.

Receiver Operating Characteristic (ROC) digunakan untuk mengevaluasi kinerja model klasifikasi biner dengan menampilkan hubungan antara Tingkat Positif Sejati (True Positive Rate / TPR) dan Tingkat Positif Palsu (False Positive Rate / FPR) pada berbagai tingkat klasifikasi. Kurva Receiver Operating Characteristic (ROC) digunakan untuk mengevaluasi kinerja model klasifikasi biner dengan menampilkan hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai tingkat klasifikasi. Operating Characteristic (ROC) digunakan untuk mengevaluasi kinerja model klasifikasi biner dengan menampilkan hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada

berbagai tingkat klasifikasi . Kurva ini menggambarkan kemampuan model untuk membedakan antara pengguna yang churn dan non – churn berdasarkan kurva menggambarkan ini kemampuan model kemampuan model untuk membedakan antara pelanggan yang akan berhenti berlangganan dan pelanggan yang tidak akan berhenti berlangganan berdasarkan satu nilai ambang batas tunggal, membedakan antara pelanggan yang berhenti berlangganan dan pelanggan yang tidak berhenti berlangganan berdasarkan satu nilai ambang batas tunggal. Daerah Nilai Area Under the Curve (AUC) Nilai Under the Curve (AUC) mewakili kinerja keseluruhan model dalam membedakan antar kelas, mewakili kinerja keseluruhan model dalam membedakan antar kelas. AUC berkisar dari 0 hingga 1, dengan 1 menunjukkan klasifikasi AUC yang ketat dan 0,5 menunjukkan lingkungan berkisar dari 0 hingga 1, dengan 1 menunjukkan klasifikasi yang ketat dan 0,5 menunjukkan lingkungan kerja yang berkelanjutan, kerja yang berkelanjutan. Dalam studi ini, model mencapai AUC sebesar 0,534, yang menunjukkan bahwa kemampuan model untuk membedakan antara pengguna yang akan berhenti berlangganan dan pengguna yang tidak akan berhenti berlangganan telah meningkat secara signifikan. Dalam penelitian ini, model tersebut mencapai AUC sebesar 0,534, yang menunjukkan bahwa kemampuan model untuk membedakan antara pengguna yang akan berhenti berlangganan dan pengguna yang tidak akan berhenti berlangganan telah meningkat secara signifikan. meningkat pesat. Angka tersebut hanya sedikit lebih tinggi daripada klasifikasi acak, yang menunjukkan bahwa model tersebut tidak mampu mengoptimalkan tingkat churn. nomor adalah hini hanya sedikit lebih tinggi darilebih tinggi acak klasifikasi acak, yang menunjukkan bahwa model tersebut tidak mampu mengoptimalkan klasifikasi, menunjukkan bahwa model tersebut tidak mampu mengoptimalkan tingkat perputaran, tingkat pergantian, Akibatnya, pengembangan lebih lanjut diperlukan untuk meningkatkan kinerja prediktif. Oleh karena itu, pengembangan lebih lanjut diperlukan untuk meningkatkan kinerja prediktif. Fitur terpenting adalah jumlah waktu yang dihabiskan untuk tugas dan frekuensi interaksi pengguna, yang menunjukkan tingkat churn yang tinggi. Faktor – faktor seperti pendapatan bulanan dan pekerjaan juga memiliki peran penting.

Pentingnya Fitur yang Dianalisis

Berikut beberapa faktor terpenting yang berkontribusi terhadap hilangnya pelanggan pada platform data streaming :



Gambar 8. Fitur Untuk Prediksi Churn.

Hasil analisis pentingnya fitur dari bahwa faktor terpenting yang memengaruhi hilangnya pelanggan yang berhenti berlangganan di platform streaming adalah pengalaman pengguna, kondisi ekonomi , dan kualitas layanan. Fitur yang paling berharga adalah kemampuan untuk mengelola waktu dan meningkatkan produktivitas , yang menyiratkan bahwa interaksi pengguna yang merupakan indikator yang baik untuk pelanggan yang baik untuk mengetahui apakah pelanggan akan berhenti berlangganan. Penurunan aktivitas menonton dapat meningkatkan minat, relevansi konten, atau kualitas pengalaman pengguna. Faktor ekonomi dan demografi , faktor-faktor,seperti pendapatan bulanan dan pekerjaan, memainkan peran penting .seperti pendapatan bulanan dan pekerjaan, memainkan peran penting . - faktor seperti jumlah promosi yang digunakan, jumlah profil yang dibuat , lamanya waktu beroperasi, dan metode Pelanggan dengan pendapatan lebih rendah cenderung lebih sensitif terhadap biaya berlangganan, sementara perbedaan usia mempengaruhi preferensi konten, tingkat adopsi teknologi, dan pola penggunaan layanan. Pada tingkat churn, faktor promosi menunjukkan loyalitas yang tinggi, namun ukuran dan durasi profil mengurangi kepercayaan pelanggan terhadap layanan.

Selain itu, besarnya jumlah permintaandan kepuasan pelanggan merupakan indikator utama kualitas layanan dan kepuasan pelangganmerupakan indikator kunci kualitas layanan . Frekuensi tinggi peminjaman yang dari dan tingkat perputaran yang tinggi mengurangi kemungkinan terjadinya masalah atau gagal bayar , yang dapat meningkatkan risiko perputaran. Peminjaman dan tingkat pergantian pelanggan yang tinggi mengurangi kemungkinan masalah atau kegagalan pembayaran dapat meningkatkan risiko kehilangan pelanggan, tinggi dan tingkat perputaran yang tinggi mengurangi kemungkinan masalah atau gagal bayar, yang dapat meningkatkan risiko kebangkrutan. Variabel lain variabel seperti lain angka jumlah promosi yang digunakan, jumlah profil yang dibuat, lamanya waktu beroperasi, dan metode pembayaran semuanya berdampak pada tingkat churn .Promosi yang digunakan, jumlah profil yang dibuat, lamanya waktu beroperasi, dan metode pembayaran semuanya berdampak pada tingkat churn (berhenti berlangganan). Faktor promosi menunjukkan loyalitas yang tinggi, namun ukuran dan durasi profil mengurangi kepercayaan pelanggan terhadap layanan, layanan tersebut mengurangi kepuasan pelanggan . Praktik pembayaran yang buruk mungkin juga menjadi sumber pergantian awal, juga bisa menjadi sumber pendapatan awal.

Implikasi Platform Streaming

Identifikasi proaktif terhadap pelanggan dengan penurunan waktu menonton, skor kepuasan rendah, atau dukungan interaksi, disertai dengan retensi terarah, sangat penting untuk mengurangi churn. Bawa platform streaming harus fokus pada peningkatan kepuasan pengguna melalui konten dan pengalaman pengguna yang dipersonalisasi , memastikan strategi dan proposal penetapan harga yang kompetitif, mengoptimalkan layanan pelanggan untuk menyelesaikan masalah dengan cepat , dan memahami berbagai kebutuhan pengguna.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa customer merupakan permasalahan strategis yang signifikan bagi platform streaming digital karena berdampak langsung terhadap keberlanjutan bisnis dan pendapatan perusahaan. Penelitian ini menggunakan algoritma Random Forest untuk memprediksi churn pelanggan dengan memanfaatkan data sekunder dari Netflix Customer Churn and Engagement Dataset. Hasil evaluasi model menunjukkan bahwa kinerja Random Forest berada pada tingkat sedang, dengan nilai akurasi sebesar 53,7%, precision 56,3%, recall 63,6%, F1-score 59,7%, dan ROC-AUC sebesar 0,534. Nilai tersebut menunjukkan bahwa model hanya sedikit lebih baik dibandingkan klasifikasi acak, sehingga belum memberikan prediksi churn yang optimal.

Analisis matriks kebingungan mengindikasikan masih tingginya jumlah false positive dan false negative, yang berpotensi menimbulkan ketidaktepatan dalam penentuan strategi retensi pelanggan. Selain itu, hasil analisis feature importance menunjukkan bahwa tingkat keterlibatan pengguna, seperti durasi menonton dan frekuensi interaksi, merupakan faktor paling dominan yang memengaruhi churn. Faktor ekonomi dan seperti pendapatan bulanan, pekerjaan, serta kepuasan pelanggan, juga pengaruh yang signifikan terhadap keputusan pelanggan untuk tetap berlangganan atau berhenti menggunakan layanan. Dengan demikian, meskipun model Random Forest mampu memberikan gambaran awal mengenai pola churn pelanggan, diperlukan pengembangan lebih lanjut untuk meningkatkan akurasi dan daya diskriminasi model.

Berdasarkan kesimpulan tersebut, beberapa saran yang dapat diberikan adalah sebagai berikut. Pertama, bagi platform streaming digital, disarankan untuk memanfaatkan hasil analisis churn sebagai dasar dalam merancang strategi retensi pelanggan yang lebih proaktif, khususnya dengan memantau penurunan aktivitas menonton, tingkat kepuasan pelanggan, serta frekuensi interaksi dengan layanan dukungan. Personalisasi konten, peningkatan kualitas pengalaman pengguna, serta penyesuaian strategi harga yang lebih fleksibel bagi pelanggan dengan sensitivitas biaya tinggi dapat menjadi langkah efektif untuk menekan tingkat churn.

DAFTAR REFERENSI

- Ahmad, A., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0191-6>
- Anderson, S. (2019). *Customer churn prediction using random forests: Analysis of machine learning techniques for churn prediction in the telecom sector*. IEEE Access, 7, 60134-60149. <https://doi.org/10.1109/access.2019.2914999>
- Beeharry, Y., & Fokone, R. (2021). Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry. *Concurrency and Computation Practice and Experience*, 34(4). <https://doi.org/10.1002/cpe.6627>
- Çallı, L., & Kasim, S. (2022). Using machine learning algorithms to analyze customer churn in the Software as a Service (SaaS) industry. *Academic Platform Journal of Engineering and Smart Systems*, 10(3), 115-123. <https://doi.org/10.21541/apjess.1139862>
- Chang, V., Hall, K., Xu, Q., Amao, F., Ganatra, M., & Benson, V. (2024). Prediction of customer churn behavior in the telecommunication industry using machine learning models. *Algorithms*, 17(6), 231. <https://doi.org/10.3390/a17060231>
- Edwine, N., Wang, W., Song, W., & Ssebuggwawo, D. (2022). Detecting the risk of customer churn in telecom sector: A comparative study. *Mathematical Problems in Engineering*, 2022, 1-16. <https://doi.org/10.1155/2022/8534739>

- Hussain, F., Neelakandan, S., Geetha, B., Selvalakshmi, V., Umadevi, A., & Martinson, E. (2022). Artificial intelligence-based customer churn prediction model for business markets. *Computational Intelligence and Neuroscience*, 2022, 1-14. <https://doi.org/10.1155/2022/1703696>
- Keramati, A., Ghaneei, H., & Mirmohammadi, S. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innovation*, 2(1). <https://doi.org/10.1186/s40854-016-0029-6>
- Muneer, A., Ali, R., Alghamdi, A., Taib, S., Almaghthawi, A., & Ghaleb, E. (2022). Predicting customers churning in banking industry: A machine learning approach. *Indonesian Journal of Electrical Engineering and Computer Science*, 26(1), 539. <https://doi.org/10.11591/ijeecs.v26.i1.pp539-549>
- Thakkar, H., Desai, A., Ghosh, S., Singh, P., & Sharma, G. (2022). Clairvoyant: AdaBoost with cost-enabled cost-sensitive classifier for customer churn prediction. *Computational Intelligence and Neuroscience*, 2022, 1-11. <https://doi.org/10.1155/2022/9028580>
- Vo, N., Liu, S., Li, X., & Xu, G. (2021). Leveraging unstructured call log data for customer churn prediction. *Knowledge-Based Systems*, 212, 106586. <https://doi.org/10.1016/j.knosys.2020.106586>
- Xie, Y., Li, X., Ngai, E., & Ying, W. (2009). Customer churn prediction using improved balanced forests. *Expert Systems With Applications*, 36(3), 5445-5449. <https://doi.org/10.1016/j.eswa.2008.06.121>
- Xu, T., Ma, Y., & Kim, K. (2021). Telecom churn prediction system based on ensemble learning using feature grouping. *Applied Sciences*, 11(11), 4742. <https://doi.org/10.3390/app11114742>
- Zhou, Y., Chen, W., Sun, X., & Yang, D. (2023). Early warning of telecom enterprise customer based on ensemble learning. *PLoS ONE*, 18(10), e0292466. <https://doi.org/10.1371/journal.pone.0292466>