

Prediksi Churn Pelanggan Industri Telekomunikasi Menggunakan Metode Artificial Neural Network Berbasis Streamlit

Nilawati Ningsih^{1*}, Nur Aida Saskiya Iskandar², Shofiatur Rizqiyah³, Sudriyanto⁴

^{1,2,3,4} Teknik Informatika, Fakultas Teknik, Universitas Nurul Jadid, Indonesia

Info Artikel

Riwayat Artikel:

Diterima : **20-Agustus-2024**

Direvisi : **23-Okttober-2024**

Disetujui : **04-Desmber-2024**

Kata Kunci:

Telekomunikasi,
Churn Pelanggan,
Artificial Neural Network,
Streamlit

ABSTRAK

Perusahaan telekomunikasi menghadapi tantangan besar dalam mempertahankan pelanggan karena biaya untuk mendapatkan pelanggan baru jauh lebih tinggi daripada mempertahankan pelanggan yang sudah ada. Churn pelanggan, atau kecenderungan pelanggan untuk berhenti menggunakan layanan, dapat menyebabkan kerugian yang signifikan bagi perusahaan. Prediksi churn pelanggan menggunakan teknik Machine Learning sangat penting untuk mengatasi masalah ini. Penelitian ini menggunakan algoritma Artificial Neural Network (ANN) berbasis Streamlit untuk memprediksi customer churn pada industri telekomunikasi. Terinspirasi dari cara kerja sistem saraf manusia, ANN dapat mempelajari pola data pelanggan yang kompleks, seperti masa aktif, jenis kontrak, dan biaya bulanan, sehingga menghasilkan prediksi yang lebih akurat. Berdasarkan hasil penelitian, metode ANN berbasis Streamlit mencapai akurasi 98%, lebih tinggi dari metode sebelumnya. Namun, akurasi yang tinggi mengindikasikan adanya potensi overfitting, sehingga pengujian lebih lanjut dengan dataset yang lebih besar dan beragam diperlukan untuk memastikan generalisasi yang lebih baik. Model ini diharapkan dapat membantu perusahaan telekomunikasi dalam mengidentifikasi pelanggan yang berpotensi berpindah dan meningkatkan strategi retensi pelanggan secara efektif.

Keywords:

Telecommunication,
Customer Churn,
Artificial Neural Network,
Streamlet

ABSTRACT

Telecommunications companies face a major challenge in retaining customers as the cost of acquiring new customers is much higher than retaining existing customers. Customer churn, or the tendency of customers to stop using a service, can cause significant losses to the company. Customer churn prediction using Machine Learning techniques is crucial to address this issue. This research uses a Streamlit-based Artificial Neural Network (ANN) algorithm to predict customer churn in the telecommunications industry. Inspired by how the human nervous system works, ANN can learn complex customer data patterns, such as tenure, contract type, and monthly fee, resulting in more accurate predictions. Based on the research results, the Streamlit-based ANN method achieved 98% accuracy, higher than the previous method. However, the high accuracy indicates the potential for overfitting, so further testing with larger and more diverse datasets is needed to ensure better generalization. This model is expected to help telecommunication companies identify potentially churning customers and improve customer retention strategies effectively.

Penulis Korespondensi:

Nilawati Ningsih,
Teknik Informatika,
Universitas Nurul Jadid ,
Email: nilawatiningasih69@gmail.com

1. PENDAHULUAN

Perusahaan telekomunikasi menjadi salah satu penyedia layanan telekomunikasi seperti telepon dan juga akses komunikasi data. Tentunya layanan telekomunikasi pada sebuah perusahaan diperuntukan kepada pelanggan atau customer [1]. Pelanggan sangat penting bagi perusahaan karena pelangganlah yang menggunakan jasa atau produk yang dihasilkan oleh perusahaan sehingga tanpa adanya pelanggan tidak akan mendapatkan keuntungan [2]. Hal demikian membuat perusahaan telekomunikasi bersaing ketat untuk mendapatkan pelanggan [3]. Perusahaan perlu mengadopsi strategi yang tidak hanya menarik pelanggan baru tetapi juga memperkuat hubungan dengan pelanggan yang sudah ada. Untuk memperoleh pelanggan baru memerlukan biaya hingga 10 kali lipat lebih mahal dibandingkan biaya untuk mempertahankan pelanggan yang ada. Mahalnya untuk memperoleh pelanggan baru tentunya perusahaan akan lebih memilih mempertahankan pelanggan [4]. Dalam konteks ini, menjaga loyalitas pelanggan lama menjadi kunci utama, pelanggan lama cenderung akan lebih loyal untuk menggunakan layanan yang disediakan karena mereka sudah mempercayai layanan yang mereka gunakan. Hal tersebut membuat pelanggan lama harus diberikan layanan yang lebih baik untuk mempertahankan mereka agar tidak churn [5].

Customer Churn didefinisikan sebagai kecenderungan pelanggan untuk berhenti melakukan bisnis dengan sebuah perusahaan [2]. *Customer Churn* menyebabkan perusahaan mengalami kerugian [6]. Dalam mempertahankan customer, Perusahaan telekomunikasi membutuhkan cara untuk memprediksi resiko kapan customer akan menjadi churn. Prediksi customer churn dapat dilakukan dengan teknik Machine Learning. Machine Learning merupakan metode ekstraksi data menjadi sebuah pola informasi tertentu [1].

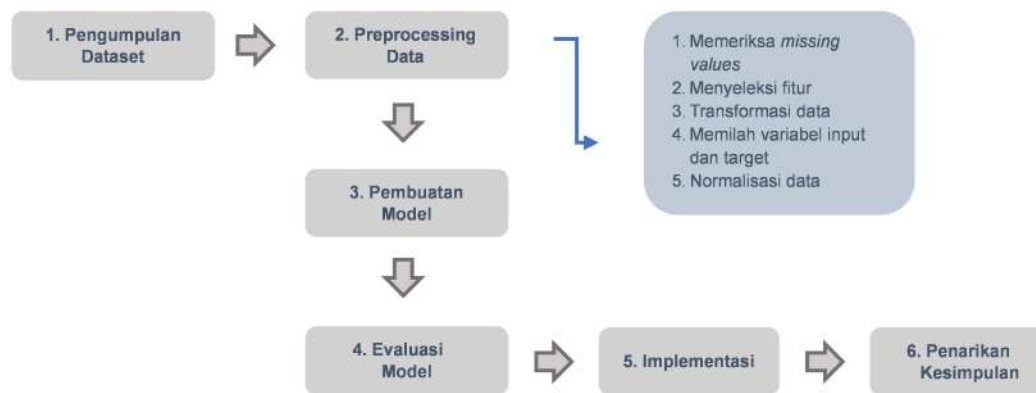
Salah satu penelitian sebelumnya yang berjudul *Prediksi Customer Churn Menggunakan Logistic Regression dan Decision Tree* disimpulkan bahwa dalam memprediksi customer churn pada sebuah perusahaan telekomunikasi model klasifikasi Logistic Regression lebih berpeluang untuk mencapai tingkat akurasi terbaik yaitu sebesar 80% dibandingkan dengan model klasifikasi Decision Tree yang hanya mencapai tingkat akurasi 72% [6]. Penelitian selanjutnya yang berjudul *Prediksi Tingkat Pelanggan Churn Pada Perusahaan Telekomunikasi Dengan Algoritma adaboost*. Kesimpulan yang didapat dari penelitiannya yaitu bahwa algoritma adaboost dapat memprediksi masalah churn lebih baik dari algoritma random forest dan xgboost serta TotalCharges adalah fitur yang paling penting dalam memprediksi churn dengan tingkat akurasi 80% [7].

Dalam penelitian ini penulis menggunakan algoritma artificial neural network berbasis streamlit untuk memprediksi pelanggan churn (berhenti menggunakan layanan) atau tidak di industri telekomunikasi serta meningkatkan akurasi prediksi dibandingkan dengan metode sebelumnya yang telah digunakan. ANN merupakan model komputasi yang terinspirasi oleh cara kerja sistem saraf manusia [8], mampu mempelajari pola kompleks dalam data historis pelanggan, seperti tenure, contract type, dan monthly charges, sehingga menghasilkan prediksi yang lebih akurat. ANN terdiri dari berbagai lapisan (layers) neuron, termasuk lapisan input, lapisan tersembunyi (hidden layers), dan lapisan output. Setiap lapisan tersembunyi terdiri dari sejumlah neuron yang menerima input, memprosesnya melalui fungsi aktivasi, dan kemudian meneruskan outputnya ke lapisan berikutnya. Penggunaan beberapa lapisan tersembunyi memungkinkan ANN untuk menangkap dan memodelkan hubungan non-linear yang kompleks dalam data. Selain itu, Streamlit adalah kerangka kerja web yang ditujukan untuk menyebarkan model dan visualisasi dengan mudah menggunakan bahasa Python, yang cepat dan minimalis tetapi juga memiliki tampilan yang cukup baik serta ramah pengguna [9]. Kombinasi ANN dan Streamlit diharapkan dapat menciptakan sistem yang tidak hanya akurat dalam memprediksi churn pelanggan tetapi juga mudah digunakan oleh pengguna.

Dengan menggunakan pendekatan ini, penelitian ini bertujuan untuk memberikan solusi yang efektif dalam mengatasi tantangan churn pelanggan di industri telekomunikasi, serta berkontribusi pada peningkatan strategi retensi pelanggan dan pengambilan keputusan yang lebih baik di dalam industri tersebut. Streamlit adalah kerangka kerja web yang ditujukan untuk menyebarkan model dan visualisasi dengan mudah menggunakan bahasa Python, yang cepat dan minimalis tetapi juga memiliki tampilan yang cukup baik serta ramah pengguna.

2. METODE PENELITIAN

Kerangka kerja penelitian merupakan tahapan prosedur yang dilakukan para peneliti untuk melakukan penelitian. Berdasarkan Gambar 1. metode penelitian dimulai dengan pengumpulan data. Setelah itu dilakukan preprocessing data pada data mentah sebelumnya agar lebih mudah pada proses pemodelan. Tahap selanjutnya adalah membuat model menggunakan *Artificial Neural Network*. Setelah model dibuat, dilakukan pengujian terhadap model untuk mengevaluasi kinerja model. Terakhir, membuat implementasi model dan penarikan kesimpulan.



Gambar 1. Kerangka Penelitian

2.1. Pengumpulan Dataset

Sumber Data: Dataset yang digunakan dalam penelitian ini adalah dataset pelanggan dari Kaggle berjudul "Telco Customer Churn" yang dapat diakses di Kaggle. Dataset ini berisi informasi tentang pelanggan perusahaan telekomunikasi, termasuk informasi demografis, layanan yang digunakan, dan informasi keuangan seperti tagihan bulanan dan total tagihan. Dataset ini terdiri dari 7043 entri dan 21 fitur, termasuk fitur target 'Churn'.

2.2. Preprocessing Data

Dataset yang didapat melalui situs Kaggle.com masih berupa data mentah dan perlu dilakukan tahap preprocessing. Tahapan preprocessing dalam penelitian ini meliputi memeriksa *missing values*, seleksi fitur, transformasi data, pemilahan variabel Input dan target, hingga yang terakhir adalah normalisasi data.

2.3. Pembuatan Model

Pada tahap pemuatan model, dipilih model algoritma Artificial Neural Network (ANN). Metode ANN dipilih karena kemampuannya dalam menangani masalah di berbagai bidang yang memiliki data-data kompleks dan non-linear [10]. Desain Arsitektur ANN yang digunakan terdiri dari beberapa lapisan penting. Pertama, input layer dengan jumlah neuron sesuai dengan jumlah fitur data. Selanjutnya, terdapat dua lapisan hidden layer masing-masing lapisan dengan 32 neuron dan 16 neuron, menggunakan fungsi aktivasi ReLU. Serta satu output layer dengan 1 neuron menggunakan fungsi aktivasi sigmoid untuk klasifikasi biner.

Model dilatih menggunakan data pelatihan dengan optimizer Adam dan loss function binary_crossentropy. Proses pelatihan dilakukan selama 100 epoch dengan batch size 32 dan 20% data pelatihan digunakan sebagai validation set.

2.4. Evaluasi Model

Evaluasi model *artificial neural networks* yang digunakan adalah *confusion matrix*. Kriteria klasifikasi dengan metode *confusion matrix* dapat dilihat pada Tabel 1.

Tabel 1. Confusion Matrix

Kelas	Diklasifikasikan sebagai Positif	Diklasifikasikan sebagai Negatif
+	True Positive (TP)	False Positive (FP)
-	True Negative (TN)	False Negative (FN)

Tabel 1 memberikan ilustrasi mengenai confusion matrix. Confusion matrix menyajikan informasi tentang jumlah prediksi benar dan salah untuk setiap kelas dalam tugas klasifikasi biner. Berikut adalah penjelasan dari setiap elemen dalam tabel tersebut:

True positive (TP) berarti model klasifikasi dengan benar memberi label jumlah tupel positif. *True Negative* (TN) berarti bahwa model klasifikasi dengan benar memberi label jumlah tupel negatif. *False positive* (FP) berarti bahwa model klasifikasi memberi label yang salah untuk jumlah tupel negatif. *False Negative* (FN) menunjukkan bahwa model klasifikasi memberi label yang salah untuk jumlah tupel positif. Akurasi adalah ukuran kinerja model klasifikasi dan merupakan persentase jumlah data yang diprediksi dengan benar dari total data [11]. Rumus menghitung akurasi model ANN ditunjukkan oleh persamaan 1.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Pengukuran *performance metrics* terdiri dari presisi, *recall* dan *f1-score* [12]. Presisi adalah rasio positif atau derajat keandalan, yaitu proporsi prediksi berlabel positif yang benar terhadap prediksi positif keseluruhan [13]. Rumus untuk menghitung presisi dapat dilihat pada Persamaan 2.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall juga dikenal sebagai *true positive rate* atau sensitivitas. *Recall* juga disebut sebagai derajat keandalan model dalam mendeteksi data berlabel positif dengan benar [13]. Rumus untuk menghitung recall dapat dilihat pada persamaan 3.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-score merangkum semua hasil perhitungan *precision* dan *recall* dengan membuat rata-rata harmonik [14]. Rumus untuk menghitung *F1-score* dapat dilihat pada persamaan 4.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

2.5. Implementasi

Implementasi sistem prediksi churn pelanggan, sebuah aplikasi web interaktif dikembangkan menggunakan Streamlit untuk memfasilitasi prediksi churn pelanggan berdasarkan input data baru dari pengguna. Pengguna dapat memasukkan data pelanggan baru melalui antarmuka yang disediakan. Aplikasi akan menggunakan model yang telah dilatih untuk memprediksi apakah pelanggan akan churn atau tidak. Model dan skaler yang telah disimpan diintegrasikan ke dalam aplikasi web untuk melakukan prediksi secara real-time berdasarkan input pengguna.

3. HASIL DAN ANALISIS

3.1. Pengumpulan Dataset

Persamaan Dataset diunduh dari Kaggle.com dengan alamat website yaitu [Telco Customer Churn \(kaggle.com\)](https://www.kaggle.com/datasets/telco-telco/customer-churn). Berikut adalah lima sampel data teratas dari dataset TelCo Customer Churn yang belum dikelola.

	customerID	gender	SeniorCitizen	Partner	Dependents	...	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
0	7590-VHVEG	Female	0	Yes	No	...	Yes	Electronic check	29.85	29.85	No
1	5575-GNVDE	Male	0	No	No	...	No	Mailed check	56.95	1889.5	No
2	3668-QPYBK	Male	0	No	No	...	Yes	Mailed check	53.85	108.15	Yes
3	7795-CFOCW	Male	0	No	No	...	No	Bank transfer (automatic)	42.30	1840.75	No
4	9237-HQITU	Female	0	No	No	...	Yes	Electronic check	70.70	151.65	Yes
5	9305-CDSKC	Female	0	No	No	...	Yes	Electronic check	99.65	820.5	Yes

6 rows × 21 columns

Gambar 1. Lima Sampel Data pada Dataset

Dataset terdiri dari 7043 baris dan 21 kolom (variable) dengan variable churn sebagai variable targetnya. Variabel-variabel dalam kolom pada dataset dijelaskan oleh Tabel 1. sebagai berikut:

Tabel 2. Deskripsi Variabel Kolom

No.	Nama Kolom	Deskripsi	Tipe Data	Isi
1	Customer ID	ID unik pelanggan	Kategorikal	-
2	Gender	Jenis Kelamin pelanggan	Kategorikal	Laki-laki, Perempuan
3	Senior Citizen	Apakah pelanggan adalah warga senior (1) atau bukan (0)	Kategorikal	1, 0
4	Partner	Apakah pelanggan memiliki pasangan	Kategorikal	Ya, Tidak
5	Dependents	apakah pelanggan memiliki tanggungan	Kategorikal	Ya, Tidak
6	Tenure	Lamanya pelanggan telah berlangganan	Numerik	(Jumlah bulan pelanggan telah berlangganan)
7	Phone Service	Apakah pelanggan memiliki layanan	Kategorikal	Ya, Tidak

telepon				
8	Multiple Lines	Apakah pelanggan memiliki beberapa saluran telepon	Kategorikal	Ya, Tidak, Tidak berlangganan telepon
9	Internet Service	Jenis layanan internet yang digunakan oleh pelanggan	Kategorikal	DSL, Fiber optic, No
10	Online Security	apakah pelanggan memiliki layanan keamanan online	Kategorikal	Ya, Tidak, Tidak berlangganan internet
11	Online Backup	apakah pelanggan memiliki layanan pencadangan online	Kategorikal	Ya, Tidak, Tidak berlangganan internet
12	Device Protection	apakah pelanggan memiliki layanan perlindungan perangkat	Kategorikal	Ya, Tidak, Tidak berlangganan internet
13	Tech Support	apakah pelanggan memiliki layanan dukungan teknis	Kategorikal	Ya, Tidak, Tidak berlangganan internet
14	Streaming TV	apakah pelanggan memiliki layanan streaming TV	Kategorikal	Ya, Tidak, Tidak berlangganan internet
15	Streaming Movies	apakah pelanggan memiliki layanan streaming film	Kategorikal	Ya, Tidak, Tidak berlangganan internet
16	Contract	Jangka waktu kontrak pelanggan	Numerik	Perbulan, Satu Tahun, Dua Tahun
17	Paperless Billing	apakah pelanggan menggunakan tagihan tanpa kertas	Kategorikal	Ya, Tidak
18	Payment Method	Metode pembayaran yang digunakan oleh pelanggan	Kategorikal	Cek elektronik, cek melalui pos, transfer bank, kartu kredit
19	Monthly Charges	Jumlah biaya bulanan yang dibayar oleh pelanggan.	Numerik	(Jumlah tagihan perbulan)
20	Total Charges	Jumlah total biaya yang telah dibayar oleh pelanggan.	Numerik	(Lama berlangganan x Tagihan bulanan)
21	Churn (Variable Target)	apakah pelanggan telah berhenti berlangganan	Kategorikal	Ya, Tidak

3.2. Preprocessing

Setelah pengumpulan dataset, selanjutnya dilakukan preprocessing data yang bertujuan bertujuan untuk mengubah data mentah menjadi format yang bersih, terstruktur, dan sesuai untuk digunakan dalam model. Berikut adalah beberapa langkah utama dalam preprocessing:

3.2.1 Pembersihan Data

Pembersihan data adalah proses mengidentifikasi dan memperbaiki atau menghapus kesalahan dan inkonsistensi dalam data. Ini mencakup langkah-langkah seperti:

- Memeriksa Missing Value

Pembersihan data dimulai dengan memeriksa apakah ada nilai yang hilang (*missing values*) di dalam dataset. *Missing values* dapat mengganggu analisis dan model prediktif, sehingga perlu diidentifikasi dan ditangani dengan tepat. langkah ini melibatkan penggunaan fungsi seperti **isnull()** atau **isna()** pada pandas untuk mendeteksi adanya *missing values* dalam setiap kolom.

Gambar 2. merupakan hasil pemeriksaan missing values pada kolom dalam dataset. Nilai nol di sebelah kanan variable kolom menunjukkan jumlah nilai yang hilang dalam kolom tersebut. Dalam hal ini, setiap kolom dalam dataset, mulai dari 'customerID' hingga 'churn', tidak memiliki nilai yang hilang (*missing values*).

```

customerID      0
gender          0
SeniorCitizen   0
Partner         0
Dependents      0
tenure          0
PhoneService    0
MultipleLines   0
InternetService 0
OnlineSecurity  0
OnlineBackup    0
DeviceProtection 0
TechSupport     0
StreamingTV     0
StreamingMovies 0
Contract        0
PaperlessBilling 0
PaymentMethod   0
MonthlyCharges  0
TotalCharges    0
Churn           0
dtype: int64

```

Gambar 2. Hasil Pemeriksaan Missing Values pada Kolom

- Seleksi Fitur

Kolom-kolom yang tidak relevan atau tidak memiliki kontribusi signifikan terhadap model harus dihapus. Contoh kolom yang mungkin dihapus adalah kolom **customerID** atau kolom yang memiliki data seragam dan tidak bervariasi. Langkah ini membantu mengurangi kompleksitas dan meningkatkan kinerja model.

Gambar 3. menunjukkan deskripsi struktur dataset yang telah dilakukan seleksi fitur dengan menghapus kolom **customerID**, dimana yang semula dataset terdiri dari 21 kolom, menjadi 20 kolom tanpa kolom **customerID**.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 20 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   gender              7043 non-null   object
 1   SeniorCitizen       7043 non-null   int64
 2   Partner             7043 non-null   object
 3   Dependents          7043 non-null   object
 4   tenure              7043 non-null   int64
 5   PhoneService        7043 non-null   object
 6   MultipleLines       7043 non-null   object
 7   InternetService     7043 non-null   object
 8   OnlineSecurity      7043 non-null   object
 9   OnlineBackup        7043 non-null   object
10   DeviceProtection    7043 non-null   object
11   TechSupport         7043 non-null   object
12   StreamingTV         7043 non-null   object
13   StreamingMovies     7043 non-null   object
14   Contract            7043 non-null   object
15   PaperlessBilling    7043 non-null   object
16   PaymentMethod       7043 non-null   object
17   MonthlyCharges      7043 non-null   float64
18   TotalCharges        7043 non-null   object
19   Churn               7043 non-null   object
dtypes: float64(1), int64(2), object(17)
memory usage: 1.1+ MB

```

Gambar 3. Deskripsi Struktur Dataset

3.2.2 Transformasi Data

Langkah transformasi data adalah mengubah data mentah menjadi format yang lebih sesuai untuk analisis atau pelatihan model machine learning. Langkah-langkah utama transformasi data meliputi:

- Pertama, mengkonversi kolom '**TotalCharges**' dari tipe data string (*object*) menjadi tipe data numerik (*float*) menggunakan '*pd.to_numeric*' dengan parameter *errors='coerce'* agar operasi matematika dan analisis numerik dapat dilakukan. Penggunaan *errors='coerce'* memastikan nilai-nilai yang tidak dapat dikonversi menjadi numerik (string kosong atau karakter *non-numeric*) diubah menjadi NaN.
- Setelah konversi, langkah selanjutnya adalah menangani *missing values* dengan mengganti nilai NaN menggunakan media dari kolom '**TotalCharges**'. Teknik ini merupakan teknik umum dalam data

preprocessing karena median tidak dipengaruhi oleh outlier dan memberikan gambaran yang lebih representative tentang distribusi data.

- Encoding Kategorikal

Melakukan label encoding untuk kolom-kolom kategorikal yang terdaftar dalam variable kolom seperti gender, Partner, Dependents, dan lainnya dikonversi menjadi bentuk numerik, misalnya 'Male' dan 'Female' dalam kolom 'gender' akan diubah menjadi 0 dan 1. Terakhir, kolom target 'Churn' dikonversi menjadi nilai numerik menggunakan 'LabelEncoder' karena model machine learning memerlukan target dalam bentuk numerik untuk melakukan klasifikasi. Hasil encoding ditunjukkan oleh Gambar 3.4 di bawah ini.

	gender	SeniorCitizen	Partner	Dependents	tenure	...	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
0	0	0	1	0	1	...	1	2	29.85	29.85	0
1	1	0	0	0	34	...	0	3	56.95	1889.50	0
2	1	0	0	0	2	...	1	3	53.85	108.15	1
3	1	0	0	0	45	...	0	0	42.30	1840.75	0
4	0	0	0	0	2	...	1	2	70.70	151.65	1

5 rows × 20 columns

Gambar 4. Dataset setelah Transformasi Data

3.2.3 Memilah Variabel Input dan Variabel Target

Langkah pertama adalah menginisialisasi daftar kolom yang digunakan sebagai fitur (input) dalam model. Variabel input disimpan dalam 'X', yang didapatkan dengan menghapus kolom 'Churn' (target) dari DataFrame asli yang nantinya akan digunakan oleh model untuk mempelajari pola dan membuat prediksi. Selanjutnya adalah memisahkan variabel target ('Churn') yang disimpan dalam 'y'. Pemisahan variabel *input* dan variabel target penting karena memungkinkan data untuk diproses secara terstruktur dan jelas.

Gambar 5. merupakan tampilan dari kolom-kolom variable input setelah dilakukan transformasi data sebelumnya menggunakan 'LabelEncoder'. Terlihat bahwa semua tipe data dalam kolom variable input telah terkonversi menjadi bentuk numerik.

	gender	SeniorCitizen	Partner	Dependents	tenure	...	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges
0	0	0	1	0	1	...	0	1	2	29.85	29.85
1	1	0	0	0	34	...	1	0	3	56.95	1889.50
2	1	0	0	0	2	...	0	1	3	53.85	108.15
3	1	0	0	0	45	...	1	0	0	42.30	1840.75
4	0	0	0	0	2	...	0	1	2	70.70	151.65

5 rows × 19 columns

Gambar 5. Tampilan Variabel Input dalam Bentuk Numerik

Gambar 6. merupakan tampilan dari kolom variable target yang telah dilakukan proses transformasi data ke dalam bentuk numerik. Target variabel Churn, yang biasanya berupa label kategori seperti "Yes" dan "No", harus diubah menjadi nilai numerik (dimana, 1 untuk "Yes" dan 0 untuk "No").

Churn
0
1
2
3
4

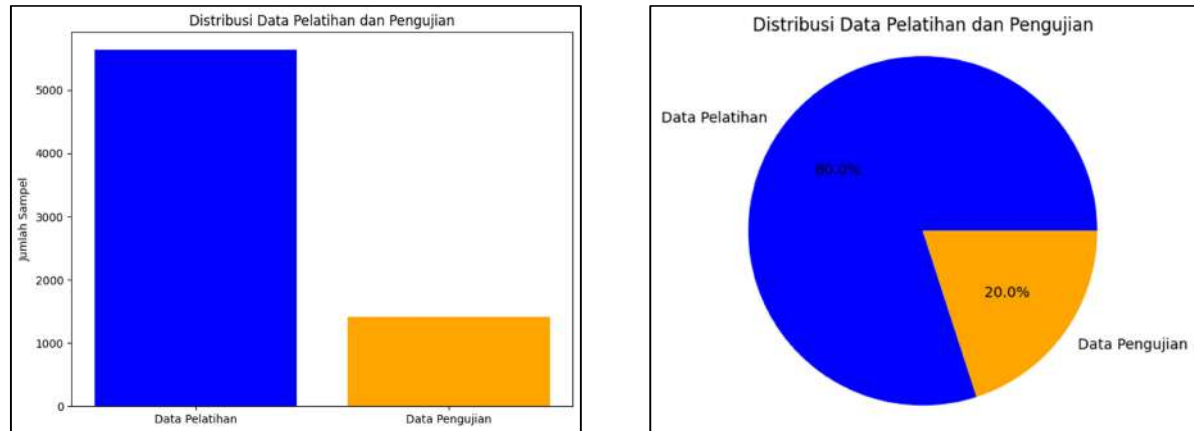
Gambar 6. Tampilan Variabel Target dalam Bentuk Numerik

3.2.4 Membagi Data Pelatihan dan Pengujian

Pada data pelatihan (*training data*), X_train memiliki ukuran (5634, 19) yang berarti terdapat 5634 sampel dengan 19 fitur masing-masing. Dan ukuran y_train sebesar (5634, 1), dimana terdapat 5634 sampel target (label) yang sesuai dengan sampel dalam X_train. Sedangkan untuk data pengujian (*testing data*), X_test memiliki ukuran (1409, 19) yang berarti terdapat 1409 sampel dengan 19 fitur masing-masing. Serta

ukuran y_{test} sebesar (1409, 1) yang berarti terdapat 1409 sampel target (label) yang sesuai dengan sampel dalam X_{test} .

Gambar 7. menunjukkan pembagian data antara pelatihan dan pengujian yang divisualisasikan ke dalam bentuk diagram batang dan pie chart. Diagram batang menunjukkan distribusi jumlah sampel dalam kedua subset data. Dan pie chart menggambarkan proporsi data pelatihan dan pengujian dalam bentuk presentase.



Gambar 7. Diagram Batang dan Pie Chart

3.2.5 Normalisasi Fitur

Data yang digunakan dalam machine learning sering kali memiliki fitur-fitur dengan skala yang berbeda. Untuk memastikan setiap fitur memiliki pengaruh yang seimbang terhadap model, normalisasi diperlukan. Pada kode tersebut, normalisasi dilakukan menggunakan **'StandardScaler'** dari pustaka **'sklearn.preprocessing'**.

3.3. Membuat Model

Model jaringan saraf tiruan (Artificial Neural Network, ANN) untuk klasifikasi biner dibangun dan dilatih menggunakan TensorFlow dan Keras. Model diinisialisasi dengan objek Sequential dan terdiri dari dua lapisan: lapisan pertama dengan 2 hidden layer masing-masing 32 dan 16 neuron menggunakan aktivasi ReLU untuk belajar pola kompleks, dan lapisan output dengan 1 neuron menggunakan aktivasi sigmoid untuk menghasilkan probabilitas antara 0 dan 1. Model dikompilasi menggunakan optimizer Adam dan fungsi kerugian binary crossentropy, serta metrik akurasi untuk evaluasi performa. Pelatihan dilakukan dengan data latih dalam batch sebesar 32 selama 100 epoch.

Setelah pelatihan, model dievaluasi menggunakan data uji untuk mengukur performa dengan nilai kerugian dan akurasi. Hasilnya menunjukkan model dapat mempelajari hubungan antara fitur input dan label kelas, memberikan prediksi probabilitas yang akurat untuk klasifikasi biner.

3.4. Evaluasi Model

Setelah model dilatih, selanjutnya adalah melakukan evaluasi kinerja model. Gambar 8. merupakan output dari hasil evaluasi model klasifikasi yang menunjukkan metrik precision, recall, dan f1-score untuk masing-masing kelas dalam dataset, serta matrik akurasi keseluruhan.

	precision	recall	f1-score	support
0	1.00	0.97	0.99	1036
1	0.93	1.00	0.97	373
accuracy			0.98	1409
macro avg	0.97	0.99	0.98	1409
weighted avg	0.98	0.98	0.98	1409

Gambar 8. Hasil Evaluasi Model

Model memiliki presisi sangat tinggi dalam memprediksi kelas 0 (1.00) dan cukup baik dalam memprediksi kelas 1 (0.93), meskipun masih ada sedikit kesalahan. Recall untuk kelas 1 adalah sempurna (1.00), sementara untuk kelas 0 sedikit lebih rendah (0.97), menunjukkan model sangat efektif dalam mengidentifikasi sampel kelas 1. Skor F1 yang tinggi (0.99 untuk kelas 0 dan 0.97 untuk kelas 1) menandakan keseimbangan baik antara presisi dan recall.

Akurasi keseluruhan model sangat tinggi, yaitu 98%, menunjukkan sebagian besar prediksi model adalah benar. Rata-rata makro dan rata-rata tertimbang menunjukkan kinerja model konsisten di seluruh kelas, meskipun ada perbedaan kecil dalam distribusi kelas. Secara keseluruhan, model menunjukkan kinerja yang sangat baik dalam memprediksi kedua kelas.

3.5. Implementasi

Dari analisis variabel menggunakan model *Artificial Neural Network* (ANN), diperoleh bahwa sistem prediksi churn pelanggan telekomunikasi berhasil diimplementasikan secara efektif. Model ANN terbukti mampu memprediksi dengan akurat kemungkinan churn pelanggan. Gambar 9 menunjukkan tampilan aplikasi web yang dikembangkan menggunakan Streamlit, yang menampilkan antarmuka untuk memasukkan data pelanggan dan hasil prediksi churn yang diperoleh.

Gambar 9. Tampilan Implementasi Hasil

4. KESIMPULAN

Berdasarkan hasil dan pembahasan, prediksi churn pelanggan telekomunikasi menggunakan metode Artificial Neural Network (ANN) berbasis Streamlit menunjukkan akurasi yang sangat tinggi, mencapai 98%. Pencapaian ini secara signifikan lebih baik dibandingkan dengan metode penelitian sebelumnya, menunjukkan bahwa model ANN ini sangat efektif dalam mengidentifikasi pelanggan yang berpotensi churn. Keakuratan yang tinggi ini memungkinkan perusahaan telekomunikasi untuk mengambil langkah-langkah proaktif dalam meningkatkan retensi pelanggan, yang pada gilirannya dapat mengurangi biaya operasional dan meningkatkan profitabilitas. Namun demikian, akurasi yang sangat tinggi juga dapat menjadi indikasi potensi overfitting, di mana model terlalu sesuai dengan data latih dan kurang mampu menggeneralisasi pada data baru. Untuk mengatasi kekhawatiran ini, pengujian lebih lanjut diperlukan dengan menggunakan dataset yang lebih besar dan lebih beragam. Hal ini penting untuk memastikan bahwa model dapat beradaptasi dengan berbagai pola data dan tetap memberikan prediksi yang akurat dalam situasi yang berbeda.

REFERENSI

- [1] M. A. Iskandar and U. Latifa, "Website Prediksi Customer Churn Untuk Mempertahankan Pelanggan Pada Perusahaan Telekomunikasi," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 2, pp. 1308–1316, 2023.
- [2] Kaharudin, M. G. Pradana, and Kusri, "PREDIKSI CUSTOMER CHURN PERUSAHAAN TELEKOMUNIKASI MENGGUNAKAN NAÏVE BAYES DAN K-NEAREST NEIGHBOR," *JURNAL INFORMASI INTERAKTIF*, vol. 4, no. 3, pp. 165–171, Sep. 2019.
- [3] A. Wicaksono, A. Anita, and T. N. Padilah, "Uji Performa Teknik Klasifikasi untuk Memprediksi Customer Churn," *Bianglala Informatika*, vol. 9, no. 1, pp. 37–45, 2021.
- [4] M. Arifin, "IG-KNN untuk prediksi customer churn telekomunikasi," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 6, no. 1, pp. 1–10, 2015.

- [5] 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.
- [6] A. M. N. Aziz, A. Mauludin, R. D. Hafiz, V. A. Sintalana, and A. A. Rismayadi, "Prediksi Customer Churn Menggunakan Logistic Regression dan Decission Tree," *eProsiding Teknik Informatika (PROTEKTIF)*, vol. 4, no. 1, pp. 11–19, 2023.
- [7] I. M. Latief, A. Subekti, and W. Gata, "Prediksi Tingkat Pelanggan Churn Pada Perusahaan Telekomunikasi Dengan Algoritma Adaboost," *Jurnal Informatika*, vol. 21, no. 1, pp. 34–43, 2021.
- [8] G. Thiodorus, A. Prasetya, L. A. Ardhani, and N. Yudistira, "Klasifikasi citra makanan/non makanan menggunakan metode Transfer Learning dengan model Residual Network," *Teknologi: Jurnal Ilmiah Sistem Informasi*, vol. 11, no. 2, pp. 74–83, 2021.
- [9] W. Hastomo, N. Aini, A. S. B. Karno, and L. M. R. Rere, "Metode Pembelajaran Mesin untuk Memprediksi Emisi Manure Management," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 11, no. 2, 2022.
- [10] W. Husada, D. Prayogo, C. F. Thamrin, and R. Herdjijono, "Optimasi Desain Penampang Struktur Rangka Batang Baja Berbasis Reliabilitas Menggunakan Symbiotic Organisms Search dan Artificial Neural Network," Petra Christian University, 2021.
- [11] S. Katoch, V. Singh, and U. S. Tiwary, "Indian Sign Language recognition system using SURF with SVM and CNN," *Array*, vol. 14, p. 100141, 2022.
- [12] X. Xiong, S. Hu, D. Sun, S. Hao, H. Li, and G. Lin, "Detection of false data injection attack in power information physical system based on SVM–GAB algorithm," *Energy Reports*, vol. 8, pp. 1156–1164, 2022.
- [13] A. Luque, A. Carrasco, A. Mart'ın, and A. de Las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit*, vol. 91, pp. 216–231, 2019.