

## Perbandingan Random Forest dan Convolutional Neural Network dalam Memprediksi Peralihan Pelanggan

Dewa Adji Kusuma <sup>(1)\*</sup>, Atika Ratna Dewi <sup>(2)</sup>, Andreas Rony Wijaya <sup>(3)</sup>

<sup>1,2</sup> Departemen Teknik Informatika, Institut Teknologi Telkom Purwokerto, Purwokerto, Indonesia

<sup>3</sup> Departemen Statistika, Universitas Sebelas Maret, Surakarta, Indonesia

e-mail : {dewaadji12, andreasronywijaya}@gmail.com, atika@ittelkom-pwt.ac.id.

\* Penulis korespondensi.

Artikel ini diajukan 21 Mei 2024, direvisi 19 Agustus 2024, diterima 20 Agustus 2024, dan dipublikasikan 31 Mei 2025.

### Abstract

The rapid growth of the telecommunications industry has increased competition among companies for customers. As a result, customers often switch to other services or terminate their subscriptions. Retaining customers is very important as it is 10 times cheaper than acquiring new customers. This study compares Random Forest (RF) and Convolutional Neural Network (CNN) algorithms in predicting customer switching, using Correlation-based Feature Selection (CFS) and Recursive Feature Elimination (RFE) for data partitioning. Model evaluation using Confusion Matrix and Area Under Curve (AUC). The evaluation results show that the performance of CNN models with optimization parameters is superior. Using the CFS dataset, the test data evaluation results yielded an accuracy of 98%, AUC of 0.96, precision of 99%, recall of 92%, and F1-score of 96%. The best tuning result for CNN is achieved with three combinations of filter and kernel sizes {[64, 7], [32, 3], [16, 2]} and a pool size of 2. A limitation of this research is determining how to compare the two algorithms being evaluated effectively. Both use different approaches, namely Supervised Learning and Deep Learning.

**Keywords:** CNN, Customer Churn, Data Mining, Prediction, Random Forest

### Abstrak

Pertumbuhan industri telekomunikasi yang pesat telah meningkatkan persaingan antar perusahaan untuk mendapatkan pelanggan. Akibatnya, pelanggan sering berpindah layanan atau menghentikan langganan. Mempertahankan pelanggan sangat penting karena 10 kali lebih murah daripada mengakuisisi pelanggan baru. Penelitian ini membandingkan algoritma Random Forest (RF) dan Convolutional Neural Network (CNN) dalam memprediksi perpindahan pelanggan, menggunakan Correlation-based Feature Selection (CFS) dan Recursive Feature Elimination (RFE) untuk partisi data. Evaluasi model menggunakan confusion matrix dan Area Under Curve (AUC). Hasil evaluasi menunjukkan performa model CNN dengan parameter optimasi lebih unggul. Menggunakan dataset CFS, hasil evaluasi data test mendapatkan akurasi sebesar 98%, AUC 0,96, precision 99%, recall 92%, dan F1-score 96%. Hasil tuning terbaik untuk CNN adalah 3 kombinasi filter dan kernel {[64&7], [32&3], [16&2]} dan pool\_size=2. Keterbatasan penelitian ini adalah bagaimana cara menangani kedua algoritma yang dibandingkan. Keduanya menggunakan pendekatan yang berbeda, yakni supervised learning dan deep learning.

**Kata Kunci:** CNN, Data Mining, Peralihan Pelanggan, Prediksi, Random Forest

### 1. PENDAHULUAN

Globalisasi dan perkembangan industri telekomunikasi yang pesat menyebabkan jumlah operator di pasar meningkat secara signifikan, menciptakan persaingan yang ketat (Lalwani et al., 2022). Menurut survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), pengguna internet di Indonesia pada periode 2022–2023 mencapai 215,63 juta orang. Selain itu, Badan Pusat Statistik (BPS) mencatat bahwa pada tahun 2021 terdapat 611 perusahaan penyedia layanan internet di Indonesia. Banyaknya perusahaan kompetitor di sektor telekomunikasi memaksa perusahaan untuk bersaing melalui strategi pemasaran yang efektif. Dampak persaingan ini adalah terjadinya perpindahan pelanggan atau *customer churn*, yaitu kondisi di mana pelanggan



Artikel ini didistribusikan mengikuti lisensi Atribusi-NonKomersial CC BY-NC sebagaimana tercantum pada <https://creativecommons.org/licenses/by-nc/4.0/>.

berhenti menggunakan layanan suatu operator karena berbagai alasan. Fenomena ini menjadi masalah serius karena pelanggan dianggap sebagai aset paling berharga oleh perusahaan, sehingga berbagai upaya dilakukan untuk mempertahankan pelanggan agar tetap berlangganan (de Lima Lemos et al., 2022; Wicaksono et al., 2021; Wu, 2024).

Mempertahankan pelanggan yang sudah ada terbukti lebih hemat, dengan biaya hingga 10 kali lebih murah dibandingkan upaya menarik pelanggan baru melalui strategi pemasaran (Gabhane et al., 2022; Ribeiro et al., 2024). Oleh karena itu, prediksi yang akurat mengenai peralihan pelanggan sangat penting untuk mendukung strategi loyalitas, perencanaan pemasaran yang efektif, serta menghasilkan penghematan biaya signifikan bagi penyedia layanan (Husein & Harahap, 2021; Li et al., 2024). Pelanggan menjadi sumber utama pendapatan bagi perusahaan telekomunikasi (AL-Najjar et al., 2022; Jain et al., 2021; Suryana et al., 2021). Dengan kemampuan memprediksi pelanggan yang berpotensi berpindah, perusahaan dapat melakukan antisipasi melalui kebijakan yang sudah dirancang, misalnya dengan memberikan layanan khusus atau insentif agar pelanggan tetap setia. Tujuan utama prediksi peralihan pelanggan adalah membantu menetapkan strategi retensi pelanggan, sekaligus mengidentifikasi sinyal awal peralihan sehingga perusahaan dapat melakukan tindakan preventif (Artha et al., 2022; Lalwani et al., 2022; Muthmainah & Cholil, 2022).

Penelitian terkait peralihan pelanggan umumnya fokus pada upaya memenangkan pelanggan di industri telekomunikasi, di mana pelanggan sangat berpengaruh terhadap pendapatan dan kapitalisasi pasar. Misalnya, penelitian Suryana et al. (2021) menggunakan metode sampling Synthetic Minority Over-sampling Technique (SMOTE) dan Boosting untuk memprediksi pelanggan yang berpindah layanan, dengan algoritma Random Forest menunjukkan akurasi unggul sebesar 89,19%. Husein & Harahap (2021) membandingkan lima model klasifikasi, yaitu K-Nearest Neighbors, Logistic Regression, Linear SVC, Random Tree, dan Random Forest, dengan hasil terbaik pada Random Forest dan K-Nearest Neighbors masing-masing mencapai akurasi 86% dan 84%. Studi tersebut juga menemukan bahwa tingkat peralihan pelanggan wanita lebih tinggi dibanding pria, dan faktor usia berpengaruh signifikan terhadap *churn*.

Penelitian Rahman & Kumar (2020) membandingkan algoritma KNN, SVM, Decision Tree, dan Random Forest, dan menemukan bahwa Random Forest dengan *oversampling* memberikan akurasi terbaik. Gabhane et al. (2022) menggunakan teknologi *deep learning* seperti Artificial Neural Network (ANN) dan Convolutional Neural Network (CNN) untuk mendekripsi *churn*, dengan CNN menunjukkan presisi lebih tinggi sebesar 97,78%. Ojo (2024) membangun tiga model prediksi *churn*, dua model Multi-layer Perceptron (MLP) dan satu CNN, dengan CNN mencapai akurasi 89%, lebih baik dibanding MLP yang hanya 80%. Yahaya et al. (2021) menggabungkan ANN dengan filter Genetic Algorithm (GA) dan K-Means untuk memprediksi *churn*, menemukan bahwa pemfilteran *noise* meningkatkan kinerja model, walaupun data tidak seimbang tetap mempengaruhi hasil.

Meskipun algoritma Random Forest dan CNN terbukti unggul, terdapat beberapa kekurangan dalam penelitian sebelumnya. Pertama, belum ada studi yang membandingkan performa kedua algoritma ini secara langsung, khususnya dalam konteks data perusahaan telekomunikasi. Kedua, karakteristik data yang berbeda yang memengaruhi kinerja masing-masing algoritma belum banyak diperhatikan.

Berdasarkan tinjauan tersebut, algoritma Random Forest dan CNN dipilih dalam penelitian ini karena memiliki performa klasifikasi dan prediksi *churn* yang baik. Random Forest efektif mengelola data dengan fitur beragam dan cenderung mengatasi *overfitting*, serta menyediakan informasi penting terkait kontribusi fitur dalam klasifikasi. Sementara itu, CNN mampu mengekstraksi fitur penting secara otomatis dari data berstruktur kompleks dan menangkap pola tersembunyi yang sulit dideteksi oleh algoritma tradisional. Evaluasi model menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, dan Area Under Curve (AUC) diharapkan memberikan wawasan terbaik mengenai algoritma yang cocok untuk prediksi *customer churn* pada perusahaan telekomunikasi.

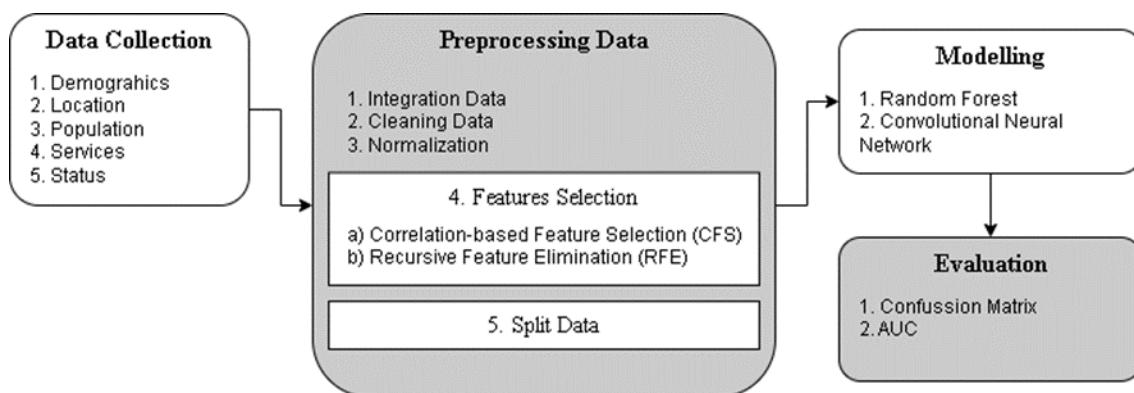


Tujuan penelitian ini adalah menganalisis dan menentukan algoritma terbaik antara Random Forest dan CNN dalam memprediksi *customer churn*. Kedua algoritma dipilih karena memiliki akurasi tinggi dan stabil menurut studi sebelumnya. Meskipun keduanya menggunakan pendekatan yang berbeda—*supervised learning* dan *deep learning*—hasil yang diperoleh dapat menjadi referensi penting untuk prediksi *churn* nyata, sehingga perusahaan dapat mengambil langkah antisipasi untuk mempertahankan pelanggan.

## 2. METODE PENELITIAN

### 2.1 Tahapan Penelitian

Untuk membantu proses penelitian, dibuat tahapan atau alur penelitian agar lebih teratur dan sistematis. Fokus utamanya adalah mengidentifikasi dan merencanakan proses yang perlu dilakukan untuk mencapai tujuan dalam penelitian. Selain itu diagram dapat digunakan sebagai alat evaluasi untuk mengevaluasi efektivitas dan efisiensi pelaksanaan penelitian. Gambar 1 menjelaskan alur penelitian yang dilakukan dalam penelitian ini.



Gambar 1 Alur Penelitian

Pengumpulan data dilakukan melalui laman *community IBM*, data yang diambil adalah data tentang *Telco customer churn* dari perusahaan telekomunikasi yang menyediakan layanan internet dan telepon rumah. *Dataset customer churn* berformat *zip* dengan beberapa *file* di dalamnya mencakup demografis, *location*, populasi, layanan, dan status. Tiap *file* memiliki jumlah baris 7.044 dengan jumlah kolom yang berbeda-beda. Semua *file* digabung menjadi satu dengan *customer\_id* sebagai kolom pembandingnya.

Selanjutnya tahapan *preprocessing data*, proses ini bertujuan untuk membersihkan, mengatur, dan mengubah data agar dapat diolah menggunakan model algoritma. Pada *intergration* data dilakukan penggabungan data menjadi satu agar akurat dan lengkap. Hasil akhir yang didapat adalah sebuah tabel dengan 40 kolom dan 7.044 baris. *Cleaning data* untuk membersihkan data dari *noise*, *missing data*, dan data yang tidak relevan untuk menjaga kualitas *dataset* dari bias. *Normalization* digunakan untuk mengubah data menjadi bentuk yang seragam dan standar.

*Features selection* digunakan untuk seleksi variabel yang akan digunakan pada model algoritma. Teknik yang digunakan pada proses ini adalah menggunakan Correlation-based Feature Selection (CFS) dan Recursive Feature Elimination (RFE). Algoritma yang digunakan pada proses RFE adalah Decision Tree. Algoritma tersebut dipilih karena hasil *feature* yang tereliminasi dinilai berbeda dibandingkan hasil CFS. *Split data* untuk membagi *dataset* menjadi data latih, data uji, dan data *test*. Pada proses ini menggunakan tiga komposisi pembagian data yaitu 60/40, 70/30, dan 80/20. Tujuan dari menggunakan tiga komposisi pembagian data adalah untuk menguji sensitivitas algoritma terhadap perbedaan data yang akan diuji, dan menganalisis optimalisasi rasio antara data latih dan uji. Penggunaan *multiple split ratios* dapat memberikan pemahaman yang lebih komprehensif tentang hasil evaluasi model.



Tahap *modelling* menggunakan algoritma Random Forest dan CNN berdasarkan dua *dataset* hasil *feature selection* serta menggunakan CFS dan RFE. Random Forest memiliki kemampuan untuk mengklasifikasi dengan menggabungkan beberapa pohon keputusan independen. CNN memiliki kemampuan dalam melihat pola-pola khusus dalam *dataset* dengan mengubahnya ke dalam bentuk matriks seperti data *sequence*, kemudian melakukan konvolusi dan *pooling* untuk mengidentifikasi fitur.

Random Forest menggunakan beberapa pohon keputusan, sehingga setiap pohon bergantung pada hasil nilai vektor acak yang dipilih secara individual menggunakan distribusi yang sama untuk semua pohon keputusan (Rahman & Kumar, 2020; Singh et al., 2021). Random Forest dimulai dengan menghitung nilai *entropy* sebagai penentu ketidakmurnian atribut dan *information gain* (Sandag, 2020). Rumus dari *entropy* dan *information gain* dinyatakan pada Pers. (1) dan (2).

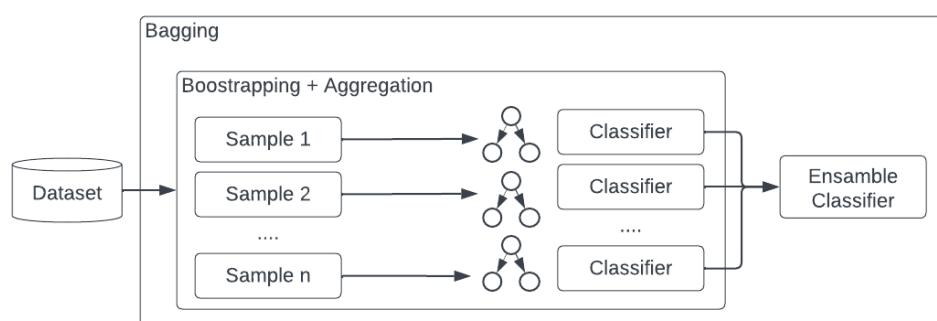
$$\text{Entropy}(Y) = - \sum_i p(c|Y) \log^2 p(c|Y) \quad (1)$$

$$\text{Inf Gain}(Y, a) = (Y) - \sum_i \nu \varepsilon \text{Values}(a) \frac{|Y_v|}{|Y_a|} (Y_v) \quad (2)$$

*Entropy*  $Y$  merupakan himpunan kasus, dan  $p(c|Y)$  adalah proporsi data dalam  $Y$  yang termasuk dalam kelas  $c$ . Formula ini digunakan untuk mengukur tingkat ketidakpastian dalam distribusi kelas yang tidak terdapat dalam himpunan data. Nilai *information gain* dari atribut  $a$  terhadap himpunan data  $Y$  dihitung dengan mengurangi nilai *entropy* total  $Y$  dengan rata-rata *entropy* dari setiap subset  $Y_v$ , yang dibagi dengan nilai dalam atribut  $a$ . Dalam hal ini,  $Y_v$  adalah subhimpunan dari  $Y$  yang memiliki nilai tertentu pada atribut  $a$ , dan  $Y_a$  adalah seluruh data yang memiliki nilai untuk atribut tersebut. Nilai  $\frac{|Y_v|}{|Y_a|}$  adalah proporsi dari masing-masing *subset* terhadap keseluruhan data yang memiliki atribut  $a$ .

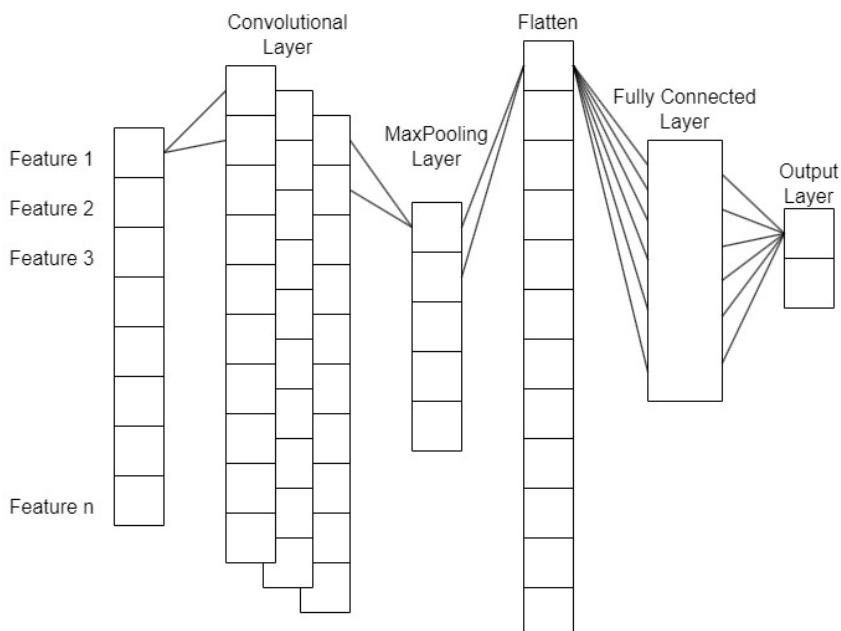
Algoritma Random Forest dimulai dengan melakukan *bootstrapping* pada *dataset*, yang menghasilkan beberapa sampel dengan sistem *sample with replacement*. Setelah itu, setiap sampel digunakan untuk membangun pohon keputusan. Hasil prediksi dari tiap pohon keputusan digabung secara *aggregation* untuk mendapatkan *ensamble classifier*. Alur algoritma Random Forest divisualisasikan pada Gambar 2.

Convolutional Neural Network (CNN) merupakan algoritma pengembangan dari Multi-layer Perceptron (MLP) yang dirancang untuk mengolah data dalam bentuk *grid* (Mawaddah et al., 2021). CNN termasuk dalam metode *supervised learning* dengan kategori *deep learning* yang dapat melakukan klasifikasi pada *dataset* berlabel. CNN memiliki beberapa *layer* yang terdiri dari *convolutional layer*, *non-linearity layer*, *pooling layer*, dan *fully connected layer* (Gabhane et al., 2022). Pada Gambar 3 akan menggambarkan ilustrasi arsitektur CNN yang diusulkan. Formula *convolution layer*, aktivasi ReLU, aktivasi *sigmoid*, *maxpooling layer*, dan *output layer* di tunjukkan pada Pers. (3) sampai (7).



**Gambar 2 Alur Algoritma Random Forest**





Gambar 3 Arsitektur Convolutional Neural Network

Bobot untuk setiap lapisan dalam CNN diinisialisasi secara acak. Untuk mengoptimalkan *hyperparameter* CNN, dilakukan pencarian *grid* pada model. Model terdiri dari lapisan *convolution layer*, menggunakan fungsi aktivasi ReLU untuk memecah linearitas, dan fungsi aktivasi *sigmoid* untuk mendapatkan *output* dengan rentang (0,1). Kemudian, pada proses *pooling* akan digunakan *MaxPooling layer* sebagai metode yang paling umum digunakan. Selanjutnya hasil dari *MaxPooling layer* akan digunakan pada proses *Flatten* dan *Fully Connected Layer* untuk mendapatkan hasil *output* akhir. Algoritma CNN lebih sering digunakan dalam melakukan klasifikasi gambar, namun CNN juga dapat diterapkan pada prediksi satu dimensi, sehingga CNN dapat digunakan untuk melakukan prediksi peralihan pelanggan.

$$output = \left\lfloor \frac{Input - Kernel Size + 2 \times Padding}{Stride} \right\rfloor + 1 \quad (3)$$

$$f_{cov}(x) = \max(0, x) \quad (4)$$

$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

$$Output = \left\lfloor \frac{Input Shape}{Pool Size} \right\rfloor \quad (6)$$

$$Output = Activ \left( \sum_{i=1}^{Input Size} Input_1 \times Weight_i + Bias \right) \quad (7)$$

Metode evaluasi yang digunakan adalah *confusion matrix* dan Area Under Curve (AUC). *Confusion matrix* adalah tabel yang memberikan gambaran dari performa sebuah model algoritma dengan spesifik. Setiap baris merepresentasikan kelas aktual data dan kolom merepresentasikan kelas prediksi data (Saputro & Sari, 2020). Tabel 1 menjelaskan mengenai *confusion matrix* dengan formula dari *confusion matrix* ditunjukkan pada Pers. (8) sampai (11). Pada persamaan tersebut, TP menjelaskan jumlah banyaknya data aktual kelas positif dan model memprediksi positif. Selanjutnya, TN menjelaskan jumlah banyaknya data aktual kelas negatif dan model memprediksi negatif. Lainnya, FP menjelaskan jumlah banyaknya data aktual kelas negatif namun model memprediksi positif. Sedangkan FN menjelaskan jumlah banyaknya data aktual kelas positif namun model memprediksi negatif.



**Tabel 1 Confusion Matrix**

	<i>Predict Negatif</i>	<i>Predict Positive</i>
<i>Actual Negatif</i>	True Negatif (TN)	False Positive (FP)
<i>Actual Positive</i>	False Negatif (FN)	True Positive (TP)

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$Presisi = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$F - measure = \frac{(Precision \times Recall)}{(Precision + Recall)} \times 2 \quad (11)$$

Akurasi adalah perbandingan antara jumlah data benar positif dan negatif dibagi dengan data keseluruhan. Presisi adalah hasil dari perbandingan antara prediksi benar positif dibagi dengan keseluruhan hasil prediksi benar positif. Recall adalah rasio prediksi benar positif dibagi dengan semua data yang benar positif. Skor F1 didapat dari hasil perkalian antara *precision* dan *recall* dibagi hasil penjumlahan dari *precision* dan *recall* kemudian dikali 2.

AUC mengukur hubungan True Positif Rate (TPR) dan False Positive Rate (FPR) pada berbagai *threshold* yang berbeda. Semakin besar nilai AUC, semakin baik performa model algoritma dalam membedakan kelas positif dan negatif. Nilai AUC berada di antara 0 hingga 1, di mana nilai 0 menunjukkan kinerja model buruk dan nilai 1 menunjukkan kinerja model baik dalam membedakan kelas positif dan negatif.

### 3. HASIL DAN PEMBAHASAN

Sebelum melakukan klasifikasi, *dataset* perlu dilakukan *preprocessing* agar data dapat diolah menggunakan model algoritma. Proses seleksi fitur menggunakan dua metode yaitu CFS dan RFE, artinya model Random Forest dan CNN akan digunakan pada kedua *dataset output* dari proses seleksi fitur. Kemudian dilakukan pembagian data seperti yang sudah dijelaskan sebelumnya. Proses *modelling* menggunakan parameter optimasi. Tabel 2 menampilkan secara ringkas parameter yang akan di optimasi. Parameter yang akan di optimasi digunakan pada proses *training* dan validasi model. Komposisi parameter beserta akurasi yang dihasilkan, dibandingkan dengan hasil menggunakan parameter *default* untuk menemukan model terbaik. Tabel 3 menampilkan hasil analisis evaluasi model menggunakan data validasi.

**Tabel 2 Parameter yang di Optimasi**

RF		CNN	
Parameter	Values	Parameter	Values
n_estimators	[50, 100, 200]	conv1_filter & kernel	[128, 64] & [7, 8]
max_depth	[0, 10, 20, 30]	conv2_filter & kernel	[64, 32] & [3, 4]
min_samples_split	[2, 5, 10]	conv3_filter & kernel	[32, 16] & [1, 2]
min_samples_leaf	[1, 2, 4]	pool_size	[2, 3]
		learning_rate	[0,01, 0,001]

Berdasarkan Tabel 3, model CNN *tuning* pada *dataset* CFS (80/20) memiliki hasil evaluasi paling baik. Evaluasi yang didapatkan adalah akurasi 99%, AUC 0,98, *precision* 100%, *recall* sebesar 97%, dan *F1-score* 98%. Model CNN tersebut digunakan pada proses klasifikasi, dan algoritma CNN *tuning* (80/20) dengan pembagian data CFS didapatkan sebagai model akhir yang memiliki performa paling baik. Selanjutnya adalah melakukan klasifikasi dengan menggunakan data *test*.



Setelah dilakukan klasifikasi, hasil akhirnya dievaluasi dan ditinjau kembali performanya. Untuk melihat hasil klasifikasi secara keseluruhan, dapat dilihat pada *confusion matrix* yang ditampilkan pada Tabel 4.

**Tabel 3 Hasil Evaluasi Model Terbaik**

Pembagian Data	Classifier	Akurasi	AUC	Precision (1)	Recall (1)	F1-score (1)
CFS (60/40)	RF	98,3%	0,97	99%	95%	97%
	RF Tuning	98,2%	0,97	99%	95%	97%
	CNN Tuning	98,5%	0,98	97%	98%	97%
CFS (70/30)	RF	98,7%	0,98	99%	97%	98%
	RF Tuning	98,8%	0,98	99%	97%	98%
	CNN Tuning	98,5%	0,98	97%	98%	97%
CFS (80/20)	RF	98,7%	0,97	99%	96%	98%
	RF Tuning	98,2%	0,97	99%	94%	97%
	CNN Tuning	99%	0,98	100%	96%	98%
RFE (60/40)	RF	98,3%	0,97	99%	95%	97%
	RF Tuning	98,3%	0,97	99%	95%	97%
	CNN Tuning	98,4%	0,97	98%	96%	97%
RFE (70/30)	RF	98,7%	0,98	99%	96%	98%
	RF Tuning	98,7%	0,98	99%	96%	98%
	CNN Tuning	98,4%	0,98	97%	97%	97%
RFE (80/20)	RF	98,4%	0,97	99%	95%	97%
	RF Tuning	98,5%	0,97	99%	95%	97%
	CNN Tuning	99%	0,97	98%	98%	99%

Dari total 167 data yang masuk ke dalam kategori peralihan pelanggan, model gagal dalam mengklasifikasi 13 data. Sedangkan dari total 538 data yang masuk ke dalam kategori tidak beralih, model gagal mengklasifikasi 1 data. Secara umum model menunjukkan hasil klasifikasi yang baik. Gambar 4 menjelaskan mengenai hasil evaluasi model pada data *test* yang sudah dilakukan. Model melakukan klasifikasi dengan baik, akurasi yang didapatkan sebesar 98% dengan AUC 0,96. *precision* yang didapatkan para proses klasifikasi sebesar 99%, *recall* 92% dan *F1-score* 96%.

**Tabel 4 Confusion Matrix**

	<b>Predict Negatif</b>	<b>Predict Positive</b>
<b>Actual Negatif</b>	537	1
<b>Actual Positive</b>	13	154

98.01418439716312

Confusion Matrix:

```
[[537  1]
 [ 13 154]]
```

	precision	recall	f1-score	support
0	0.98	1.00	0.99	538
1	0.99	0.92	0.96	167
accuracy			0.98	705
macro avg	0.98	0.96	0.97	705
weighted avg	0.98	0.98	0.98	705

AUC Score : 0.9601484762816375

**Gambar 4 Hasil Evaluasi Proses Klasifikasi**



#### 4. KESIMPULAN

Perkiraan yang akurat terkait perpindahan pelanggan memiliki dampak signifikan sebagai bagian dari strategi retensi pelanggan yang efektif. Dengan mampu memprediksi calon pelanggan yang berpotensi beralih dari layanan, perusahaan dapat mengurangi pengeluaran hingga 10 kali lipat dibandingkan dengan upaya mencari pelanggan baru. Melalui penelitian ini, model CNN dengan parameter optimasi {'conv1\_filter & kernel': 128 & 8, 'conv2\_filter & kernel': 32 & 3, 'conv3\_filter & kernel': 32 & 1, 'pool\_size': 3, 'learning\_rate': 0,001} lebih baik dibandingkan RF dan RF *Hyperparameter*. Secara spesifik, skema pembagian CFS mendapatkan hasil evaluasi yang paling baik jika dibandingkan dengan skema RFE. Akurasi sebesar 99%, AUC 0,98, *precision* 100%, *recall* 96%, dan *F1-score* 98% pada data *validation*. Pembagian *dataset* yang direkomendasikan yaitu 80/20. Pada data *test* mendapatkan akurasi sebesar 98%, AUC 0,96, *precision* 99%, *recall* 92%, dan *F1-score* 96%. Hasil ini menegaskan bahwa algoritma CNN lebih unggul dalam menangkap pola kompleks pada data pelanggan telekomunikasi, menjadikannya pilihan yang lebih baik untuk prediksi *customer churn*. Selain itu, skema pembagian data 80/20 terbukti memberikan hasil evaluasi yang optimal. Implikasi praktis dari hasil yang didapat menunjukkan bahwa penerapan model prediksi *churn* yang akurat dapat memperkuat strategi retensi pelanggan dan memberikan perusahaan keunggulan kompetitif dalam pasar yang semakin kompetitif.

Berdasarkan temuan tersebut, rekomendasi untuk penelitian selanjutnya adalah menggunakan *dataset* yang lebih besar dan beragam. Mengeksplorasi algoritma lain seperti algoritma *boosting*, serta dapat meningkatkan *epoch* pada CNN 1D dengan arsitektur yang lebih kompleks untuk memaksimalkan hasil evaluasi. Penerapan model prediksi *churn* yang akurat tidak hanya akan meningkatkan strategi retensi pelanggan, tetapi juga memberikan perusahaan keunggulan kompetitif yang signifikan dalam pasar yang semakin kompetitif. Dengan langkah-langkah ini, diharapkan model prediksi *churn* dapat mencapai akurasi yang lebih tinggi dan lebih aplikatif dalam berbagai kondisi operasional perusahaan.

#### DAFTAR PUSTAKA

- AL-Najjar, D., Al-Rousan, N., & AL-Najjar, H. (2022). Machine Learning to Develop Credit Card Customer Churn Prediction. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(4), 1529–1542. <https://doi.org/10.3390/jtaer17040077>
- Artha, B., Zahara, I., Bahri, & Permata Sari, N. (2022). Customer Retention: A Literature Review. *Social Science Studies*, 2(1), 030–045. <https://doi.org/10.47153/sss21.2952022>
- de Lima Lemos, R. A., Silva, T. C., & Tabak, B. M. (2022). Propension to Customer Churn in a Financial Institution: A Machine Learning Approach. *Neural Computing and Applications*, 34(14), 11751–11768. <https://doi.org/10.1007/s00521-022-07067-x>
- Gabhane, M. D., Suriya, A., & Kishor, S. B. (2022). Churn Prediction in Telecommunication Business Using CNN and ANN. *Journal of Positive School Psychology*, 2022(4), 4672–4680. <https://jurnalppw.com/index.php/jpsp/article/view/4158>
- Husein, A. M., & Harahap, M. (2021). Pendekatan Data Science untuk Menemukan Churn Pelanggan pada Sector Perbankan dengan Machine Learning. *Data Sciences Indonesia (DSI)*, 1(1), 8–13. <https://doi.org/10.47709/dsi.v1i1.1169>
- Jain, H., Khunteta, A., & Srivastava, S. (2021). Telecom Churn Prediction and Used Techniques, Datasets and Performance Measures: A Review. *Telecommunication Systems*, 76(4), 613–630. <https://doi.org/10.1007/s11235-020-00727-0>
- Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer Churn Prediction System: A Machine Learning Approach. *Computing*, 104(2), 271–294. <https://doi.org/10.1007/s00607-021-00908-y>
- Li, A., Yang, T., Zhan, X., Shi, Y., & Li, H. (2024). Utilizing Data Science and AI for Customer Churn Prediction in Marketing. *Journal of Theory and Practice of Engineering Science*, 4(05), 72–79. [https://doi.org/10.53469/jtpes.2024.04\(05\).10](https://doi.org/10.53469/jtpes.2024.04(05).10)
- Mawaddah, U., Armanto, H., & Setyati, E. (2021). Prediksi Karakteristik Personal Menggunakan Analisis Tanda Tangan dengan Menggunakan Metode Convolutional Neural Network



- (CNN). *Antivirus: Jurnal Ilmiah Teknik Informatika*, 15(1), 123–133. <https://doi.org/10.35457/antivirus.v15i1.1526>
- Muthmainah, & Cholil, M. (2022). Faktor yang Mempegaruhi Perilaku Peralihan Pelanggan: Peran Mediasi Kepuasan dan Kepercayaan Pelanggan. *Jurnal Riset Bisnis dan Investasi*, 7(3), 125–136. <https://doi.org/10.35313/jrbi.v7i3.3437>
- Ojo, A. K. (2024). Predicting Customer Churn in Telecommunication Industry Using Convolutional Neural Network Model. *SSRN Electronic Journal*, 22(3), 54–59. <https://doi.org/10.2139/ssrn.4983685>
- Rahman, M., & Kumar, V. (2020). Machine Learning Based Customer Churn Prediction in Banking. *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 1196–1201. <https://doi.org/10.1109/ICECA49313.2020.9297529>
- Ribeiro, H., Barbosa, B., Moreira, A. C., & Rodrigues, R. G. (2024). Determinants of Churn in Telecommunication Services: A Systematic Literature Review. *Management Review Quarterly*, 74(3), 1327–1364. <https://doi.org/10.1007/s11301-023-00335-7>
- Sandag, G. A. (2020). Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest. *CogITO Smart Journal*, 6(2), 167–178. <https://doi.org/10.31154/cogito.v6i2.270.167-178>
- Saputro, I. W., & Sari, B. W. (2020). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, 6(1), 1–11. <https://doi.org/10.24076/citec.2019v6i1.178>
- Singh, U., Rizwan, M., Alaraj, M., & Alsaidan, I. (2021). A Machine Learning-Based Gradient Boosting Regression Approach for Wind Power Production Forecasting: A Step towards Smart Grid Environments. *Energies*, 14(16), Article ID: 5196. <https://doi.org/10.3390/en14165196>
- Suryana, N., Pratiwi, P., & Prasetyo, R. T. (2021). Penanganan Ketidakseimbangan Data pada Prediksi Customer Churn Menggunakan Kombinasi SMOTE dan Boosting. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 6(1), 31–37. <https://doi.org/10.31294/ijcit.v6i1.9545>
- Wicaksono, A., Anita, A., & Padilah, T. N. (2021). Uji Performa Teknik Klasifikasi untuk Memprediksi Customer Churn. *Bianglala Informatika*, 9(1), 37–45. <https://doi.org/10.31294/bi.v9i1.9992>
- Wu, Y. (2024). Predicting Customer Churn in a Telecommunications Company Using Machine Learning. *Proceedings of the 7th International Conference on Economic Management and Green Development*, 771–783. [https://doi.org/10.1007/978-981-97-0523-8\\_72](https://doi.org/10.1007/978-981-97-0523-8_72)
- Yahaya, R., Abisoye, O. A., & Bashir, S. A. (2021). An Enhanced Bank Customers Churn Prediction Model Using a Hybrid Genetic Algorithm and K-Means Filter and Artificial Neural Network. *2020 IEEE 2nd International Conference on Cyberspac (CYBER NIGERIA)*, 52–58. <https://doi.org/10.1109/CYBERNIGERIA51635.2021.9428805>

