JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 10, No. 2, MEI, 2025, Pp. 186 – 194
ISSN: 2527 – 5836 (print) | 2528 – 0074 (online)

Perbandingan Random Forest dan Convolutional Neural Network dalam Memprediksi Peralihan Pelanggan

Dewa Adji Kusuma [(1)*], Atika Ratna Dewi [(2)], Andreas Rony Wijaya [(3)]
[1,2] Teknik Informatika, Institut Teknologi Telkom Purwokerto, Purwokerto, Indonesia
[3] 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 services or terminate 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 obtained an accuracy of 98%, AUC 0,96, precision 99%, recall 92%, and f1-score 96%. The best tuning result for CNN is 3 combinations of filter and kernel {[64&7], [32&3], [16&2]} and pool_size=2. The limitation of this research is how to handle the two algorithms being compared. 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 masif, membuat jumlah operator terus meningkat di pasar secara signifikan dan menciptakan sebuah persaingan (Lalwani et al., 2022). Menurut data survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), pengguna internet di Indonesia pada periode 2022-2023 mencapai 215,63 juta pengguna. Kemudian, menurut Badan Pusat Statistik (BPS) bahwa pada tahun 2021 terdapat 611 perusahaan penyedia layanan internet di Indonesia. Banyaknya perusahaan kompetitor dalam segi bisnis telekomunikasi membuat perusahaan tersebut harus berkompetisi dalam strategi pemasaran. Dampak dari persaingan ini menciptakan sebuah kondisi di mana pelanggan berpindah atau tidak lagi menggunakan layanan dari suatu operator karena alasan tertentu. Fenomena ini biasa dikenali dengan istilah peralihan pelanggan atau *customer churn*. Tentunya hal ini akan menjadi

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

masalah karena perusahaan menganggap pelanggan sebagai sumber daya paling berharga, sehingga perusahaan akan berupaya dengan berbagai macam cara agar pelanggan tetap melakukan pembelian tanpa berhenti (de Lima Lemos et al., 2022; Wicaksono et al., 2021).

Mempertahankan pelanggan agar tetap berlangganan layanan perusahaan dapat menghemat pengeluaran hingga 10 kali lebih murah dibandingkan dengan usaha mendapatkan pelanggan baru melalui strategi pemasaran (Gabhane et al., 2022). Memperkirakan peralihan pelanggan dengan akurat secara efektif dapat mendukung strategi loyalitas pelanggan, merencanakan *marketing* ekonomi dan menghasilkan penghematan yang signifikan bagi penyedia layanan (Husein & Harahap, 2021; Li et al., 2024). Pelanggan merupakan hal yang penting bagi perusahaan telekomunikasi karena merupakan pemasukan utama untuk perusahaan (AL-Najjar et al., 2022; Suryana et al., 2021). Sebenarnya, jika perusahaan dapat memprediksi pelanggan yang akan beralih dari produk perusahaan, maka perusahaan dapat mengantisipasi hal tersebut dengan kebijakan yang sudah disepakati. Perusahaan sebagai penyedia layanan dapat melakukan analisis data dan melakukan prediksi untuk mengetahui perilaku pelanggan serta melakukan prediksi kapan pelanggan akan berpindah ke layanan kompetitor. Hal ini dapat menjadi bahan pertimbangan perusahaan untuk melakukan antisipasi dengan memberikan pelayanan khusus atau layanan tambahan untuk mempertahankan pelanggan. Tujuan utama dari memprediksi peralihan pelanggan adalah membantu menetapkan strategi retensi pelanggan (Lalwani et al., 2022). Menetapkan strategi untuk melacak pelanggan yang loyal merupakan keharusan karena model prediksi peralihan pelanggan bertujuan untuk mengidentifikasi sinyal peralihan pelanggan awal dan mencoba memprediksi pelanggan yang akan beralih dari layanan perusahaan (Lalwani et al., 2022; Muthmainah & Cholil, 2022).

Penelitian yang terkait peralihan pelanggan, umumnya membahas mengenai kompetisi untuk menarik pelanggan. Pada perusahaan telekomunikasi, pelanggan berpengaruh sangat besar terkait pemasukan dan kapitalisasi pasar. Pada penelitian Suryana et al. (2021) dengan tujuan memprediksi pelanggan yang berpindah layanan menggunakan optimasi pada level data, melalui metode sampling Synthetic Minority Over Sampling. Hasil dari penelitian ini menunjukkan algoritma Random Forest lebih unggul setelah dioptimalisasi menggunakan SMOTE dan Boosting dengan hasil akurasi 89,19%. Penelitian lain Husein & Harahap (2021) dengan tujuan untuk melakukan analisis dan prediksi peralihan pelanggan menggunakan lima model yang berbeda yaitu Kneighbors *classifier,* Logistic Regression*,* Linear SVC, Random Tree *classifier* dan Random Forest *classifier*. Hasil dari penelitian ini menunjukan bahwa algoritma *Random Forest* dan Kneighbors *classifier* lebih baik dibandingkan algoritma lainnya dengan akurasi 86% dan 84%. Kesimpulan dari penelitian ini ialah pelanggan Wanita lebih tinggi tingkat peralihannya dibandingkan dengan pelanggan pria dan tingkatan umur sangat mempengaruhi peralihan pelanggan.

Penelitian Rahman & Kumar (2020) dengan tujuan memprediksi dan *customer churn* menggunakan algoritma KNN, SVM, Decision Tree, dan Random Forest. Hasilnya, model Random Forest setelah *oversampling* lebih baik dibandingkan dengan model lainnya dalam hal akurasi. Penelitian Gabhane et al. (2022) dengan tujuan mengidentifikasi *customer churn* menggunakan teknologi *deep learning* seperti Artificial Neural Network (ANN) dan Convolutional Neural Network (CNN). Hasilnya menunjukkan bahwa algoritma CNN lebih baik dengan presisi 97,78% dibandingkan ANN. Penelitian Ojo (2024) dengan tujuan membangun model untuk memprediksi *customer churn* menggunakan CNN. membangun model menjadi 3 model yang terdiri dari dua model Multi-layer Perceptron (MLP) dan satu model CNN. Hasilnya menunjukkan bahwa tingkat akurasi CNN lebih baik sebesar 89% dibandingkan MLP sebesar 80%. Penelitian Yahaya et al. (2021) dengan tujuan memprediksi *customer churn* menggunakan Artificial Neural Network (ANN) dengan filter Genetic Algorithm (GA) dan filter K-Means. Hasilnya menunjukan bahwa kinerja model meningkat ketika noise difilter, sementara kinerja model dipengaruhi oleh data yang tidak seimbang yang disebabkan oleh pemfilteran.

Meskipun penelitian sebelumnya sudah menunjukkan keunggulan algoritma Random Forest dan CNN dalam memprediksi peralihan pelanggan, terdapat beberapa gap yang ada. Pertama, belum

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

ada penelitian yang membandingkan performa kedua algoritma ini terutama pada algoritma dengan pendekatan yang berbeda dalam konteks data perusahaan telekomunikasi. Kemudian, penelitian sebelumnya tidak mempertimbangkan perbedaan karakteristik data yang mempengaruhi kinerja tiap algoritma.

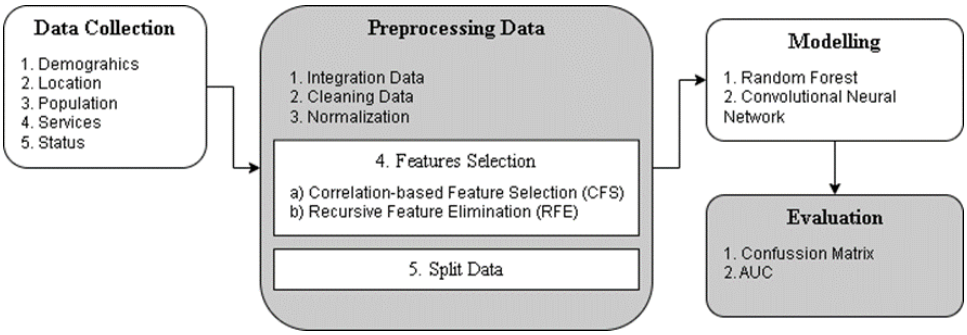
Berdasarkan penelitian terdahulu yang sudah dijelaskan sebelumnya, algoritma Random Forest dan CNN memiliki performa yang baik dalam klasifikasi dan memprediksi *customer churn*. Pemilihan kedua algoritma ini didasari pada beberapa pertimbangan. Random Forest mampu menangani data dengan fitur yang beragam yang sering ditemui pada data perusahaan umumnya. Random Forest cukup efektif dalam mengatasi *overfitting* dan dapat memberikan informasi tentang pentingnya fitur dalam proses klasifikasi. Algoritma CNN dipilih karena mampu menangani data dengan struktur yang kompleks dan mampu mengekstraksi fitur penting secara otomatis. Dalam konteks data pelanggan telekomunikasi, CNN mampu menangkap pola tersembunyi yang tidak didapatkan algoritma tradisional. Berdasarkan keunggulan dari kedua algoritma tersebut, metode evaluasi model yang digunakan adalah akurasi, *precision*, *recall*, *F1-score* dan Area Under Curve (AUC). Membandingkan kedua algoritma ini dapat memberikan wawasan baru mengenai pendekatan terbaik untuk prediksi *customer churn* dalam konteks perusahaan telekomunikasi.

Tujuan dari penelitian ini adalah untuk menganalisis dan menemukan algoritma terbaik antara algoritma Random Forest dan CNN. Kedua algoritma dipilih karena memiliki *range* akurasi yang tinggi dan stabil berdasarkan penelitian sebelumnya. Keterbatasan penelitian ini adalah cara menangani kedua algoritma yang dibandingkan. Keduanya menggunakan pendekatan yang berbeda, yakni *supervised learning* dan *deep learning*. Algoritma terbaik yang didapatkan selanjutnya dapat menjadi bahan rujukan untuk prediksi peralihan pelanggan dalam kasus nyata, sehingga perusahaan dapat melakukan 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. Langkah selanjutnya akan masuk ke dalam tahap *preprocessing*.

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

JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 10, No. 2, MEI, 2025: 186 – 194

Preprocessing data 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). Formula dari *entropy* dan *information gain* dinyatakan pada Pers. (1) dan (2).

$$Entropy(Y) = -\sum_i p(c \mid Y) \log^2 p(c \mid Y) \tag{1}$$

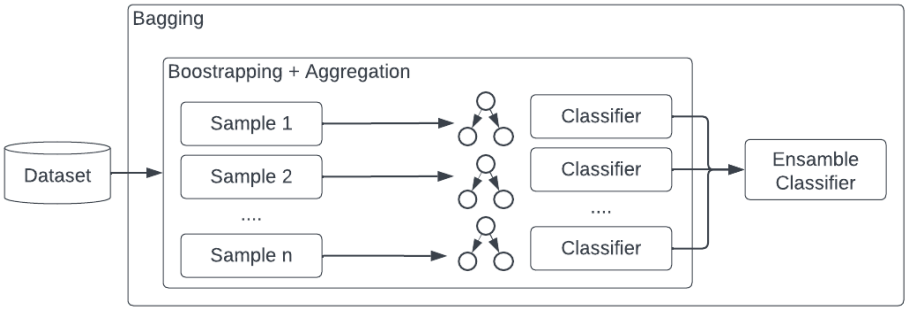
$$InfGain(Y, a) = (Y) - \sum v \varepsilon Values(a) \frac{|Y_v|}{|Y_a|} (Y_v) \tag{2}$$

Entropy Y merupakan himpunan kasus, dan $p(c \mid Y)$ 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 *boostrapping* 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

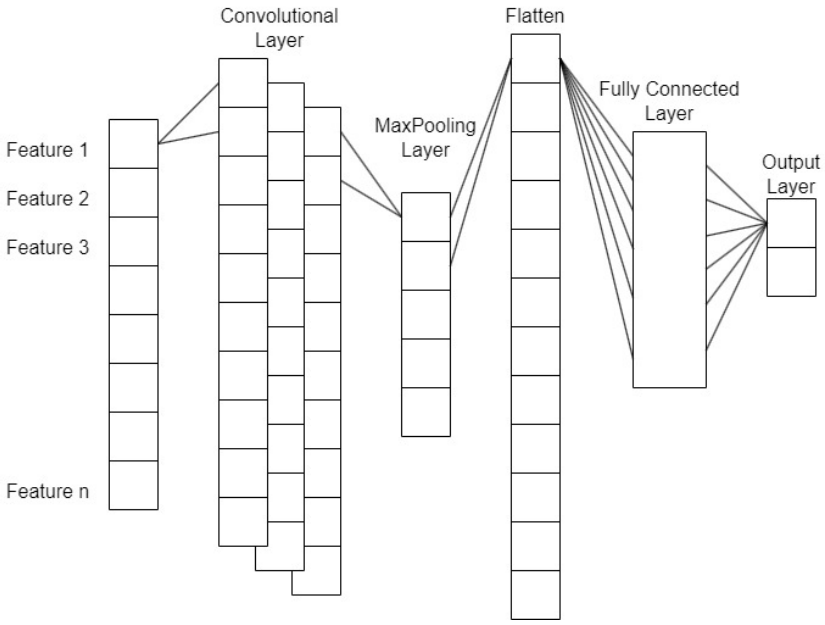
Artikel ini didistribusikan mengikuti lisensi Atribusi-NonKomersial CC BY-NC sebagaimana tercantum pada https://creativecommons.org/licenses/by-nc/4.0/.

digabung secara *aggregation* untuk mendapatkan *ensamble classifier*. Alur algoritma Random Forest divisualisasikan pada Gambar 2.



Gambar 2 Alur Algoritma Random Forest

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 illustrasi arsitektur CNN yang diusulkan. Formula *convolution layer*, aktivasi ReLU, aktivasi *sigmoid*, *maxpooling layer*, dan *output layer* di tunjukan pada Pers. (3) sampai (7).



Gambar 3 Arsitektur Convolutional Neural Network

Bobot untuk setiap lapisan dalam CNN diinisialisasi secara acak. Untuk mengoptimalisasi *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 *ouput* 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.

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

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

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

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

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

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

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*

	Predict Negatif	***Predict Positive***
Actual Negatif	*True Negatif* (TN)	*False Positive* (FP)
Actual Positive	*False Negatif* (FN)	*True Positive* (TP)

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

$$\text{Pr}esisi = \frac{TP}{TP + FP} \quad (9)$$

$$\text{Re}call = \frac{TP}{TP + FN} \quad (10)$$

$$F - measure = \frac{(\text{Pr}ecision * \text{Re}call)}{(\text{Pr}ecision + \text{Re}call)} * 2 \quad (11)$$

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.

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

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]

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%

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.

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

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. *precission* yang didapatkan para proses klasifikasi sebesar 99%, *recall* 92% dan *F1-score* 96%.

Tabel 4 *Confusion Matrix*

	Predict Negatif	***Predict Positive***
Actual Negatif	537	1
Actual Positive	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,

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

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

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

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://journalppw.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

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

Mawaddah, U., Armanto, H., & Setyati, E. (2021). Prediksi Karakteristik Personal Menggunakan Analisis Tanda Tangan dengan Mengggunakan 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

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. 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), 5196. https://doi.org/10.3390/en14165196

Suryana, N., Pratiwi, P., & Prasetio, 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

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

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