

PREDIKSI CUSTOMER CHURN PADA PERUSAHAAN TELEKOMUNIKASI MENGUNAKAN ALGORITMA C4.5 BERBASIS PARTICLE SWARM OPTIMIZATION

Moch. Rizki Kurniawan, Puspita Nurul Sabrina, Ridwan Ilyas

Teknik Informatika, Universitas Jenderal Achmad Yani

Jl. Terusan Jend. Sudirman, Cibeber, Kec. Cimahi Sel., Kota Cimahi, Jawa Barat 40531

moch.rizki@student.unjani.ac.id

ABSTRAK

Customer churn didefinisikan sebagai kecenderungan customer berhenti dan berpindah dari satu layanan kepada layanan lain pada periode tertentu, Karena ini merupakan masalah terpenting dalam dunia bisnis khususnya dalam industri telekomunikasi yang sangat mempengaruhi keuntungan perusahaan, dalam mendapatkan pelanggan baru biayanya jauh lebih tinggi dari pada mempertahankan pelanggan yang sudah ada. Memprediksi customer churn dapat membantu sektor CRM (Customer Relationship Management) pada perusahaan menentukan strategi mempertahankan pelanggan yang ada. Penelitian ini membahas mengenai prediksi customer churn menggunakan algoritma c4.5 berbasis particle swarm optimization yang digunakan untuk seleksi atribut dan optimasi pengukuran evaluasi confusion metrics. Skenario pembagian data latih 80% dan data uji 20%. Dari hasil penelitian optimasi particle swarm optimization mampu menyeleksi atribut pada algoritma C4.5, sehingga menghasilkan tingkat akurasi dalam prediksi customer churn yang lebih baik dibanding dengan menggunakan metode individual algoritma C4.5 terjadi peningkatan nilai akurasi sebesar 1.91% dengan nilai akurasi sebesar 91.55%, precision sebesar 85.21%, recall sebesar 85.0% dan f1-score sebesar 85.11%.

Kata kunci: Algoritma C4.5, Particle Swarm Optimization, telekomunikasi, customer churn.

1. PENDAHULUAN

Perkembangan teknologi telekomunikasi telah mengalami pertumbuhan yang signifikan karena permintaan konsumen terhadap layanan komunikasi dan informasi terus meningkat seiring berjalannya waktu, berdasarkan data yang diperoleh dari survei Wearesocial & Hootsuite tahun 2023, populasi global saat ini mencapai 8,01 miliar orang. Dari jumlah tersebut, terdapat sekitar 8,46 miliar individu yang menggunakan perangkat mobile, 5,16 miliar orang yang mengakses internet, dan sekitar 4,76 miliar pengguna aktif media sosial [1]. Berdasarkan hasil survey yang telah dijelaskan, terlihat bahwa kemajuan dalam komunikasi telah mengakibatkan masyarakat menjadi lebih cenderung menggunakan layanan telekomunikasi secara intensif. Potensi industri telekomunikasi masuk ke dalam persaingan yang besar karena perubahan yang cepat, liberalisasi pasar, fitur teknis baru, rencana menarik yang berbeda, persaingan di bidang telekomunikasi semakin ketat dikarenakan pelanggan memiliki kebebasan untuk memilih layanan yang mereka inginkan dari banyak penyedia layanan yang ada. Dampak dari persaingan tersebut adalah tingginya kemungkinan pelanggan berpindah layanan ke perusahaan penyedia jasa telekomunikasi lain dan akan semakin sulit mendapatkan pelanggan. Dalam lingkup bisnis, hal ini dikenal sebagai *Customer churn* [2]. *Customer churn* adalah pelanggan yang berpindah atau berhenti dari satu layanan kepada layanan lain pada periode tertentu [2]. *Customer churn* ini merupakan masalah paling krusial dalam dunia bisnis karena sangat mempengaruhi laba perusahaan, beberapa industri telekomunikasi internasional

mengalami tingkat *churn* yang cukup tinggi tahunan rata-rata sebesar 27% [3] dan mendapatkan pelanggan baru memakan biaya 10 kali lebih mahal dari pada mempertahankan pelanggan yang sudah ada [4]. Oleh karena itu, dalam usaha mempertahankan pelanggan, perlu melakukan deteksi dini untuk mengetahui potensi pelanggan yang memiliki kecenderungan untuk *churn* pada waktu tertentu. Salah satunya adalah dengan memanfaatkan teknik data mining untuk melakukan prediksi *customer churn*. Data mining merupakan teknik yang digunakan untuk menggali pola-pola dalam data. Terdapat beragam metode yang dapat digunakan untuk tujuan prediksi atau klasifikasi dalam data mining. Beberapa metode yang umum digunakan termasuk metode pohon keputusan (*decision tree*), metode bayesian, dan jaringan saraf tiruan (*neural network*). Selain itu, juga ada metode optimisasi yang berfungsi untuk mencari solusi optimal pada berbagai jenis masalah. Salah satu metode yang cukup dikenal adalah *particle swarm optimization*.

Pada penelitian [5] tentang Prediksi *churn rate* pada jasa telekomunikasi menggunakan algoritma C4.5, untuk langkah awal dalam mengidentifikasi pola persentasi perpindahan pelanggan pada perusahaan telekomunikasi. Penelitian [6] implementasi menggunakan IG-KNN gabungan algoritma seleksi fitur *information gain* dengan KNN dan Penelitian [7] implementasi menggunakan Algoritma Naïve bayes dalam memprediksi *customer churn* diantara metode data mining tersebut yang telah digunakan untuk prediksi *Customer churn* hasilnya belum mencapai nilai yang sangat baik.

Algoritma C4.5 merupakan suatu metode pembuatan pohon keputusan yang digunakan dalam analisis data dan pemodelan prediktif. Metode ini juga telah digunakan pada penelitian terdahulu [5] dalam memprediksi *customer churn* didapat hasil prediksi relatif lebih baik dari metode lainnya [6] [7]. PSO adalah algoritma optimisasi terinspirasi dari perilaku kelompok burung atau ikan yang mengarahkan partikel dalam pencarian solusi optimal melalui iterasi. PSO digunakan untuk seleksi atribut dan membantu mengurangi dimensi input yang tidak relevan dan meningkatkan performa prediksi.

Penelitian ini mengimplementasikan metode algoritma C4.5 berbasis PSO (*Particle Swarm Optimization*) dalam memprediksi *customer churn* pada perusahaan telekomunikasi PSO dipilih untuk seleksi atribut Kemudian dievaluasi menggunakan confusion matrix dan *curve ROC(AUC)* sehingga ditemukan hasil nilai terbaik dalam memprediksi *customer churn*.

2. TINJAUAN PUSTAKA

Tinjauan pustaka membahas teori-teori yang mendukung serta penelitian sebelumnya mengenai *customer churn* pada perusahaan telekomunikasi.

2.1. Data Mining

Data mining adalah gambaran umum dari isi sekelompok data sebagai tujuan untuk menemukan informasi yang berguna dalam data yang sebelumnya tidak diketahui [8]. Data mining sering digunakan untuk mengelola kumpulan data, dengan hasil membantu pengambil keputusan dalam membuat keputusan terbaik. Data mining merupakan tahapan dalam proses *Knowledge Discovery in Database* (KDD), digunakan untuk memperoleh informasi yang berguna dari data. Penggunaan pohon keputusan cara pendekatan yang paling menjanjikan dan populer. Pohon keputusan merupakan metode langsung yang efektif untuk memprediksi dan menjelaskan hubungan antara berbagai pengukuran tentang item dan nilai targetnya.

Proses tersebut dilakukan dengan Langkah-langkah berikut ini:

1. Pembersihan data, proses menghilangkan data yang bising(noise), tidak benar, dan tidak relevan dari database. seperti data yang hilang atau salah, atau kesalahan ketik sederhana.
2. Integrasi data, Proses menggabungkan data dari berbagai sumber menjadi satu tampilan terpadu untuk pengelolaan data yang efisien, wawasan yang bermakna, dan kecerdasan yang dapat ditindak lanjuti.
3. Seleksi Data, ini berdampak pada data penelitian yang dihasilkan, prosedur tersebut harus didokumentasikan secara menyeluruh. Data berdasarkan proses yang tidak lengkap atau kode yang cacat adalah contoh data yang mungkin dianggap tidak relevan (dan karenanya dibuang).

4. Transformasi data, Data diubah menjadi format yang dapat digunakan untuk pemrosesan.
5. Proses mining, metode utama untuk mengekstrak informasi berharga pada data.
6. Evaluasi pola, Untuk identifikasi pola menarik yang ditemukan berbasis pengetahuan. Hasil berupa pola tipikal dan model prediksi dan dievaluasi [9]

2.2. Customer Churn

Customer Churn merupakan keadaan di mana pelanggan memutuskan untuk beralih dari satu penyedia layanan telekomunikasi ke penyedia lain yang menawarkan jasa atau produk yang serupa. Hal ini menjadi masalah yang penting dan harus diatasi oleh perusahaan-perusahaan dalam industri ini, terutama mengingat biaya yang diperlukan untuk mendapatkan pelanggan baru memiliki tingkat kesulitan yang jauh lebih tinggi dibandingkan dengan menjaga pelanggan yang telah ada [4] Churn dapat dibagi menjadi tiga jenis:

1. Churn Aktif/Disengaja: Pelanggan dengan sengaja memutuskan untuk berhenti menggunakan layanan dan beralih ke perusahaan lain karena berbagai alasan seperti kurangnya pemahaman tentang skema layanan, ketidakmampuan perusahaan dalam menyelesaikan masalah, biaya yang terlalu tinggi, kurangnya insentif loyalitas, atau masalah privasi.
2. Churn Rotasi/Insidental: Pelanggan berhenti menggunakan layanan tanpa niat untuk beralih ke penyedia lain karena alasan seperti masalah keuangan pribadi, perubahan lokasi geografis, atau tidak lagi memerlukan produk atau layanan yang ditawarkan oleh perusahaan.
3. Churn Pasif: Perusahaan menghentikan kontrak dengan pelanggan karena pelanggan tidak dapat membayar tagihan yang ada.

Churn pelanggan dalam industri telekomunikasi adalah suatu fenomena yang tidak dapat dihindari, namun perusahaan dapat mengurangi risiko churn dengan melakukan upaya mitigasi guna mengurangi kerugian yang dialami [10]

2.3. Algoritma C4.5

Algoritma C4.5 merupakan algoritma pembelajaran yang diawasi, algoritma ini memerlukan penggunaan kumpulan data pelatihan. Sebuah sampel diklasifikasikan digunakan dalam set data pelatihan. C4.5 menganalisis kumpulan data pelatihan dan membuat pengklasifikasi yang harus mengklasifikasikan data pelatihan dan pengujian dengan benar[11].

Formula algoritma C 4.5 dapat dilihat pada formula (1) - (2)[12]

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (1)$$

Dengan penjelasan $\log_2$:

$$\log_2 = \frac{\ln(x)}{\ln(2)}$$

Dimana

1. S = Himpunan Kasus.
2. P_i = proporsi dari S_i terhadap S . S_i diperoleh dari $Gain$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Dimana

1. S = Himpunan Kasus
2. A = Atribut
3. $|S_i|$ = jumlah seluruh kasus pada partisi ke-i.
4. $|S|$ = jumlah seluruh sampel data.
5. $Entropy(S_i)$ = Jumlah kasus pada partisi ke-i

2.4. Particle Swarm Optimization

Particle Swarm Optimization (PSO) dalam sebagian besar istilah dasar termasuk dalam paradigma kecerdasan *swarm*, yang mempelajari perilaku kolektif dan karakteristik sosial dari sistem yang terorganisir, terdesentralisasi, dan kompleks yang dikenal sebagai "gerombolan". Segerombolan adalah kumpulan (populasi) individu yang bergerak yang tampaknya tidak terorganisir yang cenderung mengelompok bersama-sama sementara setiap individu tampaknya bergerak ke arah yang acak. Setiap individu dalam *swarm* memiliki kemampuan interaksi dengan individu lain atau yang disebut "agen" atau "partikel" dalam PSO, dalam metode PSO dasar (bPSO), segerombolan partikel terbang melalui ruang pencarian dimensi N , dimana posisi setiap partikel mewakili potensial solusi untuk masalah optimasi. Setiap partikel a dalam kawanan, meskipun kemampuan masing-masing agen agak dibatasi oleh seperangkat aturan tertentu. Oleh karena itu, perilaku agen dalam gerombolan sering tidak signifikan, tetapi perilaku kolektif dan sosial mereka adalah sangat penting, dalam hal itu, kecerdasan kawanan berasal dari adaptasi kolektif dan sifat stokastik dari kawanan. Motivasi utama berasal langsung dari kawanan organik di alam seperti kawanan burung, gerombolan ikan, semut koloni, dan kawanan hewan lainnya, yang menunjukkan diri yang menakjubkan organisasi dan kemampuan adaptasi kolektif/sosial [13]

Persamaan untuk perhitungan PSO untuk perpindahan dan kecepatan partikel:[14]

$$\begin{aligned} V_i(t) &= V_i(t-1) + c_{1r_1}[X_{pbest\ i} - X_i(t)] + \\ &\quad c_{2r_2}[X_{Gbest} - X_i(t)] \\ X_i(t) &= X_i(t-1) + V_i(t) \end{aligned} \quad (3)$$

Dimana:

$V_i(t)$: kecepatan individu/partikel i saat iterasi

$X_i(t)$: posisi individu/partikel i saat iterasi

c_1 dan c_2 : learning rate.

$X_{pbest\ i}$: partikel i dengan posisi terbaik saat iterasi

X_{Gbest} : posisi terbaik kelompok atau populasi

2.5. Confusion matrix

Confusion matrix adalah suatu tabel yang digunakan dalam pembelajaran mesin dan statistik untuk mengukur kinerja model klasifikasi. Ini membantu menggambarkan performa model dengan membandingkan prediksi yang dibuat oleh model dengan hasil sebenarnya dari data yang diuji. Metode uji confusion matrix digunakan untuk menentukan evaluasi metrics dapat dilihat dalam Tabel 1 [15]

Tabel 1. Confusion matrix

Aktual	Prediksi	
	positif	negatif
positif	TP	FP
negatif	FN	TN

Dimana:

- True Positive (TP) merupakan jumlah pelanggan yang benar-benar Churn, dan prediksi sebagai Churn.
- False Positive (FP) adalah jumlah pelanggan yang diidentifikasi sebagai Churn, tetapi prediksi mengidentifikasinya sebagai bukan Churn.
- False Negative (FN) adalah jumlah pelanggan yang sebenarnya bukan Churn, tetapi prediksi mengidentifikasinya sebagai Churn.
- True Negative (TN) adalah jumlah pelanggan yang benar-benar bukan Churn, dan prediksi berhasil mengidentifikasinya sebagai bukan Churn.

2.6. Kurva ROC (Reciever Operating Characteristic) dan AUC (Area Under Curve).

ROC (*Receiver Operating Characteristic*) adalah sebuah grafik yang menggambarkan hubungan antara sebuah grafik atau kurva yang digunakan untuk menilai sejauh mana performa model klasifikasi. Kurva ini terutama membantu dalam membandingkan seberapa baik model dalam mengenali kasus positif yang sebenarnya, dengan seberapa sering model memberikan prediksi yang salah dengan menganggapnya sebagai kasus positif palsu. TPR digunakan untuk mengukur kemampuan model dalam mengidentifikasi dengan benar kasus positif yang sebenarnya, sementara FPR mengukur seberapa sering

model membuat kesalahan dengan mengklasifikasikan kasus negatif sebagai positif.[16] AUC merupakan sebuah metrik numerik yang digunakan untuk menilai kualitas kinerja model berdasarkan kurva ROC. Skala nilai AUC berkisar antara 0 hingga 1, di mana semakin mendekati 1, semakin baik kemampuan model dalam memisahkan kelas yang berbed.

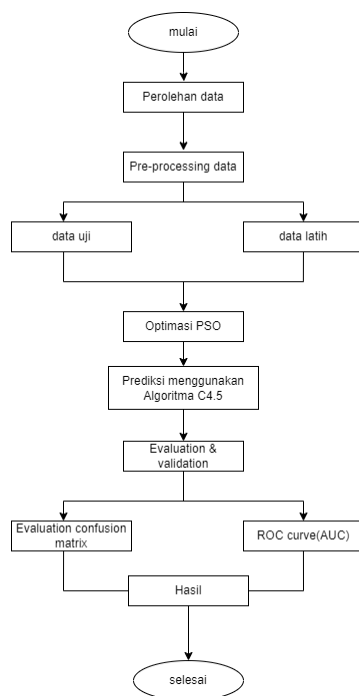
Persamaan untuk menghitung AUC: [17]

$$AUC = \frac{1}{2} \sum_{i=1}^n (x_{i+1} - x_i)(y_{i+1} - y) \quad (4)$$

Tabel 2. Kriteria nilai AUC

Nilai AUC	Interpretasi
0.9-1.00	Excellent Classification
0.8-0.9	Good Classification
0.7-0.8	Fair Classification
0.6-0.7	Poor Classification
0.5-0.6	Failure

3. METODE PENELITIAN



Gambar 1. Metode penelitian

Tahap Pertama Perolehan data, dataset yang digunakan pada penelitian ini didapatkan dan diunduh dari website *Kaggle*. Dataset yang digunakan berukuran 1.37 MB. Dataset ini memiliki 7.043 data pelanggan dan 33 atribut dengan kelas label *churn*.

Tahap Kedua yang dilakukan setelah data terkumpul adalah proses Preprocessing. Pada tahapan ini terdapat beberapa langkah yang akan dilakukan mencari dan menemukan missing value pada dataset. Transformasi data, perubahan skala pengukuran dari

data asli ke bentuk lain agar sesuai analisis variabel. Dan Normalisasi data proses mengubah data dalam skala tertentu agar memenuhi persyaratan atau aturan yang ditetapkan agar mampu memudahkan dalam proses pemodelan.

Tahap Ketiga Data dipisahkan dan dikelompokkan kedalam data training(latih) dan data testing(uji), data latih menggunakan 80% sebanyak 5634 dan data uji menggunakan 20% sebanyak 1409 dari keseluruhan dataset.

Tahap Keempat menggunakan algoritma PSO sebagai metode optimalisasi yang akan diimplementasikan. Bobot paling optimal untuk setiap atribut akan diperoleh dengan menggunakan hasil pembobotan algoritma PSO sebagai dasar pemilihan atribut yang akan digunakan dalam menentukan status customer churn.

Tahap Kelima Penerapan algoritma C4.5 menggunakan pendekatan serakah dengan membangun pohon keputusan secara top-down, di mana pemisahan dan pemilihan atribut dilakukan dengan hasil optimasi PSO sebelumnya.

Tahap terakhir Proses Evaluasi matriks konfusi digunakan untuk mengukur kinerja atau tingkat keakuratan dari suatu proses klasifikasi. Data pelatihan diklasifikasikan sebagai positif atau negatif berdasarkan prediksi yang dibuat oleh metode algoritma C4.5, dan matriks N x N untuk mengevaluasi kinerja model pada klasifikasi, N sebagai jumlah kelas target. Matriks ini untuk membandingkan nilai target aktual dengan prediksi mode pembelajaran mesin dan validasi menggunakan kurva ROC(AUC). Setelah semua tahapan dilakukan, akan mendapatkan hasil dari proses pemodelan dan evaluasi yang telah dilakukan.

4. HASIL DAN PEMBAHASAN

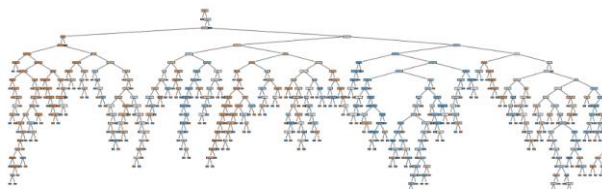
4.1. Hasil

Penelitian ini bertujuan untuk meningkatkan akurasi prediksi customer churn pada Perusahaan telekomunikasi dengan menerapkan *Particle Swarm Optimization* (PSO) dalam proses eliminasi atribut pada metode Algoritma C4.5, tidak semua dataset digunakan dalam pemodelan hanya atribut yang relevan yang telah dilakukan proses pre-processing data akan digunakan dalam proses pemodelan, di mana eksperimen akan membandingkan prediksi customer churn pada Perusahaan telekomunikasi menggunakan individual C45 dengan metode seleksi atribut menggunakan PSO. Semua eksperimen akan dijalankan pada dataset yang sudah melewati proses validasi.

4.2. Eksperimen Dan Pengujian Algoritma C4.5

Model Algoritma C4.5 diimplementasikan dengan menggunakan dataset yang terdiri dari 22 atribut yang telah dilakukan pre-processing, yang menggambarkan atribut prediksi *customer churn*. Hasil dari algoritma C4.5 berupa sebuah struktur pohon keputusan dan hasil evaluasi metrics serta kurva

ROC(AUC), Hasil pohon keputusan dapat dilihat seperti gambar dibawah ini:



Gambar 2. Pohon Keputusan C4.5

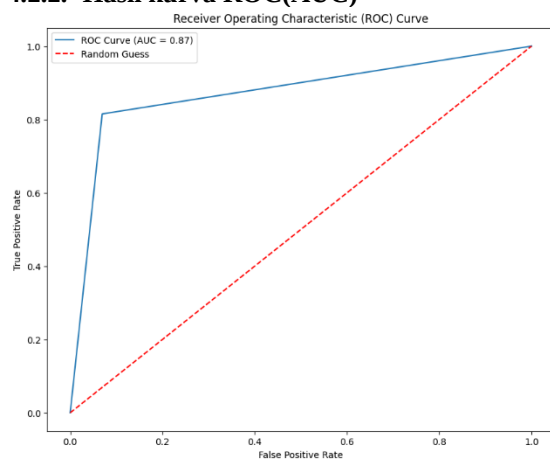
4.2.1. Hasil Evaluasi Model Dengan Confusion Matrix

Kemudian, dilakukan pengujian dan validasi dengan menggunakan confusion matrix dan kurva ROC untuk menilai akurasi dan modelnya. Hasil evaluasi confusion matrix dapat dilihat pada tabel 3.

Tabel 3. Hasil C4.5

	Nilai
Akurasi	89.64%
Presisi	82.07%
Recal	81.25%
F1-score	81.66%

4.2.2. Hasil kurva ROC(AUC)

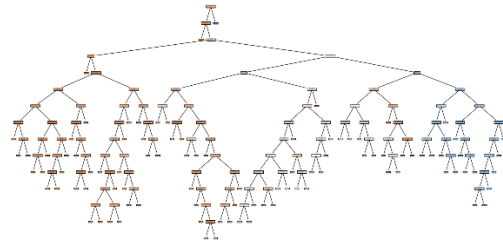


Gambar 3. Kurva ROC(AUC) C4.5

Dengan AUC sebesar 0,87 yang terlihat pada gambar di atas, dapat disimpulkan bahwa tingkat akurasi berada dalam kategori Good Classification.

4.3. Eksperimen Dan Pengujian Model Algoritma C4.5 Berbasis PSO

Selanjutnya untuk memperbaiki tingkat performance algoritma C.45 maka dilakukan optimasi menggunakan particle swarm optimization (PSO). Hasil pohon keputusan dapat dilihat seperti gambar dibawah ini:



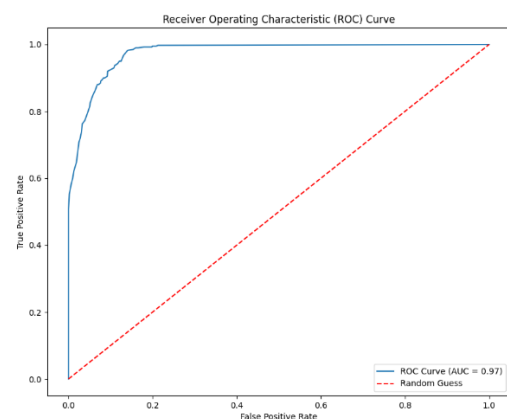
Gambar 4. Pohon Keputusan C4.5+PSO

4.3.1. Hasil Evaluasi Model Dengan Confusion Matrix

Tabel 4. Hasil C4.5 berbasis PSO

	Nilai
Akurasi	91.55%
Presisi	85.21%
Recal	85.0%
F1-score	85.11%

4.3.2. Hasil kurva ROC(AUC)



Gambar 5. Kurva ROC(AUC) C4.5 +PSO

Dengan nilai AUC sebesar 0,97 yang diperlihatkan oleh grafik di atas, ini mengindikasikan bahwa tingkat akurasi berada dalam kategori Excellent Classification.

4.4. Hasil Pengujian Model dan Perbandingan

Di bawah ini adalah tabel perbandingan tingkat ketepatan antara hasil prediksi menggunakan algoritma C4.5 yang dioptimalkan dengan PSO dan hanya menggunakan algoritma C4.5.

Tabel 5. Hasil Perbandingan C4.5 & C4.5 berbasis PSO

Perbandingan	C4.5	C4.5+PSO
Akurasi	89.64%	91.55%
Presisi	82.07%	85.21%
Recal	81.25%	85.0%
F1-score	81.66%	85.11%

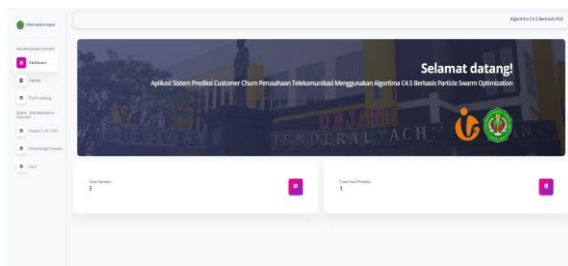
Berdasarkan table 5 diatas, penggunaan metode C4.5+PSO menghasilkan akurasi sebesar 91.55% paling tinggi pada data latih 80% dengan metrik

evaluasi lainnya yang cukup baik, seperti presisi, recall, dan f-1 score, dalam prediksi *customer churn* telekomunikasi. Sedangkan penggunaan metode C4.5 saja menghasilkan akurasi sebesar 89.64% dengan performa yang sedikit lebih rendah pada metrik evaluasi lainnya. Hal ini menunjukkan bahwa penggunaan metode C4.5+PSO dapat memberikan hasil yang lebih baik dengan peningkatan akurasi sebesar dalam prediksi *customer churn* telekomunikasi. Selain itu, dari kurva ROC, dapat dilihat bahwa metode C4.5+PSO memiliki AUC (Area Under the Curve) sebesar 0.97, sedangkan metode C4.5 memiliki AUC sebesar 0.87, yang menunjukkan tingkat keakuratan yang lebih tinggi pada metode C4.5+PSO.

4.5. Layout Hasil Rancangan Aplikasi

Hasil rancangan aplikasi sistem Prediksi Customer churn pada Perusahaan telekomunikasi menggunakan algoritma C4.5 Berbasis *Particle Swarm Optimization*.

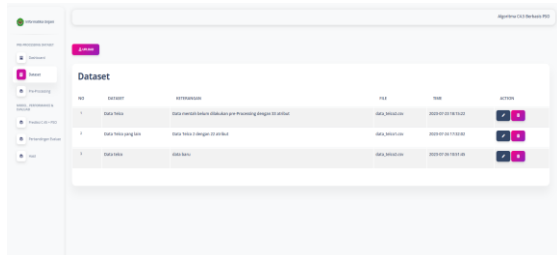
1. Halaman Dashboard



Gambar 6. Halaman dashboard sistem

Pada Gambar diatas merupakan fitur dashboard yang berisi informasi yang dibutuhkan serta beberapa proses komponen dalam memprediksi dengan algoritma C4.5 berbasis PSO.

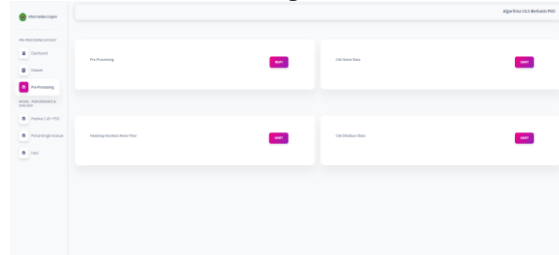
2. Halaman Dataset



Gambar 7. Halaman dataset sistem

Pada Gambar 7 merupakan fitur dataset yang berisi informasi dataset yang telah disimpan ke database serta fungsi upload file untuk menyimpan dataset, pada fitur ini juga terdapat crud yang mampu melakukan hapus dan edit data pada dataset yang telah disimpan didatabase.

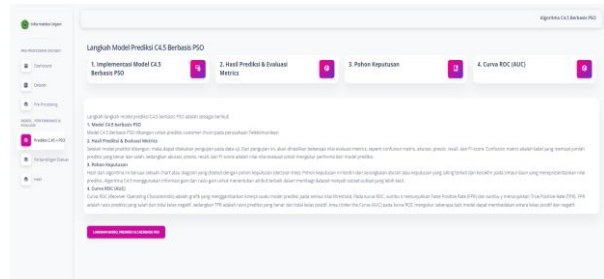
3. Halaman Pre-Processing Data



Gambar 8. Halaman Pre-Processing data sistem

Pada Gambar 8 merupakan fitur pre-processing terdapat informasi untuk melakukan proses pre-processing data pada dataset yang telah diupload atau disimpan sebelumnya, pada fitur ini dapat melakukan pre-processing data, cek noise pada data, cek distribusi data, dan membuat heatmap korelasi antar fitur pada dataset.

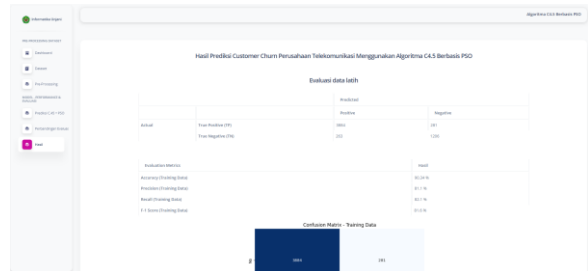
4. Halaman Prediksi Model C4.5 Berbasis PSO



Gambar 9. Halaman Model C4.5 Berbasis PSO

Pada Gambar 9 merupakan fitur prediksi model C4.5 berbasis PSO yang akan diproses dari dataset yang telah upload dan dilakukan pre-processing data.

5. Halaman Hasil



Gambar 10. Halaman Hasil sistem

Pada gambar 10 merupakan tampilan hasil halaman untuk melihat hasil yang telah diproses oleh model algoritma C4.5 berbasis PSO.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa penggunaan data mining untuk memprediksi *customer churn* dalam perusahaan telekomunikasi dengan memanfaatkan algoritma C4.5 yang dioptimalkan oleh PSO berhasil meningkatkan kinerja algoritma C4.5. Akurasi yang semula sebesar 89,64% meningkat menjadi 91,55%, menghasilkan peningkatan sebesar 1,91%. Selain itu, nilai presisi

adalah 85,21%, recall adalah 85,0%, dan f1-score adalah 85,11% saat menggunakan 80% data latih. Selain itu, didapatkan nilai AUC sebesar 0,97 dengan interpretasi *Excellent Classification*. Dapat disimpulkan penggunaan teknik optimasi *particle swarm optimization* mampu menyeleksi atribut pada algoritma C4.5. Hasilnya adalah peningkatan tingkat akurasi dalam memprediksi customer churn yang lebih baik daripada ketika hanya menggunakan individual algoritma C4.5, dengan hasil yang didapat dapat membantu perusahaan telekomunikasi untuk memprediksi *customer churn*

Saran penelitian selanjutnya mengganti dataset *customer telco* dengan dataset lain dan menggunakan pengembangan dalam peningkatan model prediksi dengan menggunakan metode-metode optimasi lainnya, optimasi yang dapat digunakan untuk meningkatkan performa prediksi churn, seperti algoritma genetika

DAFTAR PUSTAKA

- [1] simon kemp, "DIGITAL 2023: GLOBAL OVERVIEW REPORT," *dateportal.com*.
- [2] K. Coussement, S. Lessmann, and G. Verstraeten, "A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry," *Decis Support Syst*, vol. 95, pp. 27–36, Mar. 2017, doi: 10.1016/j.dss.2016.11.007.
- [3] T. Y. Fei, L. H. Shuan, L. J. Yan, G. Xiaoning, and S. W. King, "Prediction on Customer Churn in the Telecommunications Sector Using Discretization and Naïve Bayes Classifier," *Int. J. Advance Soft Compu. Appl*, vol. 9, no. 3, 2017.
- [4] S. KhakAbi, M. R. Gholamian, and M. Namvar, "Data mining applications in customer churn management," in *ISMS 2010 - UKSim/AMSS 1st International Conference on Intelligent Systems, Modelling and Simulation*, 2010, pp. 220–225. doi: 10.1109/ISMS.2010.49.
- [5] Y. Tri Utami, D. Asiah Shofiana, Y. Heningtyas, J. Ilmu Komputer FMIPA Universitas Lampung Jl Soemantri Brodjonegoro No, and B. Lampung, "Penerapan Algoritma C4.5 Untuk Prediksi Churn Rate Pengguna Jasa Telekomunikasi," 2020. doi: 10.23960%2Fkomputasi.v8i2.2647.
- [6] M. Arifin, "IG-KNN untuk prediksi customer churn telekomunikasi," *Jurnal SIMETRIS*, vol. 6, 2015, doi: 10.24176/simet.v6i1.230.
- [7] R. Novendri, R. Andreswari, and O. N. Pratiwi, "Implementasi data mining untuk memprediksi customer churn menggunakan algoritma naive bayes," 2021.
- [8] L. Rokach and O. Maimon, "Data mining with decision trees," 2014. [Online]. Available: <http://www.worldscientific.com/series/smpai>
- [9] S. Agarwal, "Data mining: Data mining concepts and techniques," in *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*, Institute of Electrical and Electronics Engineers Inc., Oct. 2014, pp. 203–207. doi: 10.1109/ICMIRA.2013.45.
- [10] N. Wayan Wardani, "Penerapan Data Mining Dalam Analytic CRM," 2021. [Online]. Available: <https://www.researchgate.net/publication/351776255>
- [11] K. Grąbczewski, "Meta-Learning in Decision Tree Induction." [Online]. Available: <http://www.springer.com/series/7092>
- [12] W. Dai and W. Ji, "A mapreduce implementation of C4.5 decision tree algorithm," *International Journal of Database Theory and Application*, vol. 7, no. 1, pp. 49–60, Jan. 2014, doi: 10.14257/ijda.2014.7.1.05.
- [13] Adaptation and Learning, "Serkan Kiranyaz Turker Ince Moncef Gabbouj Multidimensional Particle Swarm Optimization for Machine Learning and Pattern Recognition." [Online]. Available: <http://www.springer.com/series/8335>
- [14] M. H. Moradi and M. Abedini, "A combination of genetic algorithm and particle swarm optimization for optimal DG location and sizing in distribution systems," *International Journal of Electrical Power and Energy Systems*, vol. 34, no. 1, pp. 66–74, Jan. 2012, doi: 10.1016/j.ijepes.2011.08.023.
- [15] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Inf Sci (N Y)*, vol. 340–341, pp. 250–261, May 2016, doi: 10.1016/j.ins.2016.01.033.
- [16] X. Zhang, X. Li, Y. Feng, and Z. Liu, "The use of ROC and AUC in the validation of objective image fusion evaluation metrics," *Signal Processing*, vol. 115, pp. 38–48, Oct. 2015, doi: 10.1016/j.sigpro.2015.03.007.
- [17] Gorunescu and B. H. Springer, "Data Mining: Concepts, Models and Techniques," 2011. doi: 10.1007/978-3-642-19721-5_5.