PENERAPAN ALGORITMA GRID SEARCH UNTUK MEMPREDIKSI CUSTOMER CHURN PADA BANK MENGGUNAKAN PERBANDINGAN OPTIMASI DECISION TREE DAN RANDOM FOREST

Sabrina Laila Sari, Basuki Rahmat*, Kartini
Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur
Jalan Rungkut Madya, Gn. Anyar, Surabaya, Jawa Timur, Indonesia
basukirahmat.if@upnjatim.ac.id

ABSTRAK

Industri perbankan saat ini beroperasi dalam ekosistem yang sangat kompetitif, sehingga retensi nasabah menjadi sama pentingnya dengan akuisisi baru. Biaya memperoleh nasabah baru umumnya lebih tinggi dibandingkan mempertahankan nasabah yang ada, sehingga kemampuan mendeteksi potensi *customer churn* menjadi kebutuhan strategis. Perilaku *churn* dipengaruhi oleh berbagai faktor seperti demografi, aktivitas rekening, jumlah produk, dan status keaktifan nasabah. Pola kompleks tersebut sulit ditangkap dengan analisis sederhana, sehingga pendekatan *machine learning* lebih sesuai. Permasalahan utama penelitian ini adalah bagaimana mengidentifikasi nasabah berisiko *churn* menggunakan data demografi dan perilaku rekening, serta sejauh mana optimasi *hyperparameter* pada model Decision Tree dan Random Forest dapat meningkatkan performa deteksi, khususnya pada kelas minoritas. Penelitian ini bertujuan untuk membandingkan efektivitas kedua model setelah dilakukan optimasi menggunakan algoritma Grid Search, dengan fokus pada metrik akurasi, *recall*, *f1-score*, ROC-AUC, dan PR-AUC. Metode penelitian menggunakan dataset publik berisi 10.000 entri dengan fitur terstruktur, melalui tahapan pra-pemrosesan berupa encoding, normalisasi, analisis korelasi, dan pembagian data pada skenario 70:30 dan 80:20. Hasil pengujian menunjukkan bahwa optimasi *hyperparameter* meningkatkan kualitas klasifikasi. Decision Tree mencapai akurasi 0.86 namun terbatas dalam mendeteksi nasabah berisiko tinggi, sedangkan Random Forest tampil lebih unggul dengan ROC-AUC 0.838 dan PR-AUC 0.650. Secara keseluruhan, Grid Search terbukti memperbaiki performa model, dan Random Forest direkomendasikan sebagai pilihan utama untuk sistem deteksi *customer churn* di perbankan.

Kata kunci : *customer churn, grid search, decision tree, random forest, prediksi, optimasi hyperparameter*

1. PENDAHULUAN

Industri perbankan beroperasi dalam persaingan ketat, sehingga retensi nasabah menjadi sama pentingnya dengan akuisisi baru. Biaya memperoleh nasabah baru umumnya lebih tinggi dibanding mempertahankan yang sudah ada, sehingga kemampuan mendeteksi potensi *churn* menjadi kebutuhan strategis. Perilaku *churn* dipengaruhi berbagai faktor seperti demografi, aktivitas rekening, jumlah produk, dan status keaktifan. Pola kompleks ini sulit ditangkap analisis sederhana, sehingga pendekatan *machine learning* lebih sesuai.

Algoritma berbasis pohon keputusan, yakni Decision Tree dan Random Forest, dipilih karena keunggulan interpretasi dan kestabilan. Namun performa keduanya sangat bergantung pada pemilihan *hyperparameter*. Untuk itu, penelitian ini menerapkan Grid Search sebagai metode sistematis dalam menemukan konfigurasi optimal. Dataset publik berisi 10.000 entri digunakan dengan pra-pemrosesan berupa encoding, normalisasi, analisis korelasi, serta pembagian data 70:30 dan 80:20.

Hasil pengujian menunjukkan bahwa akurasi keseluruhan relatif mirip, tetapi perbedaan jelas muncul pada metrik sensitif terhadap kelas minoritas, seperti PR-AUC dan *recall*. Decision Tree mudah diinterpretasi namun kurang efektif dalam mengenali nasabah berisiko tinggi, sedangkan Random Forest lebih konsisten dengan pemisahan kelas yang stabil. Penelitian ini menegaskan bahwa optimasi *hyperparameter* melalui Grid Search meningkatkan akurasi, *recall*, dan *f1-score*, sekaligus memperkuat generalisasi. Dengan demikian, Random Forest direkomendasikan sebagai model utama, sementara Decision Tree tetap relevan untuk kebutuhan interpretabilitas.

2. TINJAUAN PUSTAKA

Pada penelitian ini dibutuhkan acuan dari penelitian terdahulu yang berkaitan dengan masalah yang akan dibahas sebgaia pembanding serta pedoman dalam memahami suatu metode yang digunakan, yaitu:

Dengan judul "Machine learning based customer churn prediction in banking". Yaitu, Penelitian ini mempromosikan eksplorasi kemungkinan churn dengan menganalisis perilaku pelanggan. Pengklasifikasi KNN, SVM, Decision Tree, dan Random Forest digunakan dalam penelitian ini. Selain itu, beberapa metode pemilihan fitur telah dilakukan untuk menemukan fitur yang lebih relevan dan untuk memverifikasi kinerja sistem. Eksperimen dilakukan pada dataset pemodelan churn dari Kaggle. Hasilnya dibandingkan untuk menemukan model yang sesuai dengan presisi dan prediktabilitas yang lebih tinggi. Hasilnya, penggunaan model Random Forest setelah oversampling lebih baik dibandingkan dengan model lain dalam hal akurasi.[1]

Dengan judul "Machine learning based telecom-customer churn prediction". Yaitu, Penelitian kami difokuskan pada penerapan beberapa teknik ini: Random Forest, SVM, Extreme Gradient Boosting (XGBoost), Ridge classifier, K-nearest neighbours (KNN) dan Deep Neural Networks. Kumpulan data yang digunakan difokuskan pada program retensi pelanggan yang mencakup berbagai bidang atribut pelanggan dan juga kolom churn pelanggan. Kami melakukan langkah-langkah pra-pemrosesan pada kumpulan data ini dan yang sama digunakan sebagai data input untuk menerapkan semua teknik. Perbandingan antara semua teknik yang disebutkan dilakukan untuk mengklasifikasikan apakah pelanggan akan churn atau tidak. Efisiensi model-model ini dieksplorasi lebih lanjut dengan melewatinya melalui pencarian grid. Oleh karena itu, disimpulkan bahwa model Random Forest bekerja paling baik untuk kasus penggunaan khusus ini dengan akurasi prediksi sebesar 90,96% pada data pengujian sebelum pencarian grid. Melalui penelitian ini, kesimpulan yang relevan dapat ditarik mengenai efektivitas model yang lebih lama tetapi lebih ringan dalam memprediksi churn pelanggan.[2]

2.1. Kerangka Teori

Kerangka teori berfungsi sebagai landasan konseptual yang mendasari penelitian ini. Penelitian ini berfokus pada penerapan algoritma dalam *machine learning* untuk memprediksi *customer churn* di sektor perbankan. Beberapa teori yang relevan dalam penelitian ini meliputi konsep churn, penerapan algoritma pembelajaran mesin, serta proses optimasi model menggunakan teknik Grid Search.

2.2. Prediksi *Customer Churn*

Prediksi *customer churn* merupakan upaya menganalisis data nasabah untuk memperkirakan potensi kehilangan pelanggan dalam periode tertentu. Bagi perbankan, hal ini krusial karena biaya akuisisi nasabah baru umumnya jauh lebih besar dibandingkan biaya mempertahankan nasabah yang sudah ada. Dengan adanya model prediksi churn, bank dapat merancang strategi retensi yang lebih tepat sasaran melalui pemahaman terhadap karakteristik serta perilaku nasabah yang memiliki risiko tinggi untuk berhenti menggunakan layanan.

2.3. Algoritma Decision Tree

Decision Tree merupakan algoritma klasifikasi yang menggunakan struktur pohon untuk memecah data berdasarkan fitur yang paling relevan. Proses pembentukan pohon dilakukan secara bertahap dan rekursif dengan memilih atribut yang memberikan pemisahan kelas terbaik, biasanya ditentukan melalui kriteria seperti *Gini Index* atau *Entropy*. Setiap simpul internal menggambarkan suatu kondisi pada fitur tertentu, sedangkan simpul daun (leaf node) menunjukkan hasil prediksi kelas.

Keunggulan utama dari Decision Tree adalah kemudahan interpretasi serta visualisasi, sehingga sering digunakan pada tahap analisis awal maupun untuk menjelaskan hasil model kepada pihak non-teknis. Namun, algoritma ini memiliki kelemahan berupa kecenderungan mengalami *overfitting*, terutama bila pohon terlalu kompleks dan tidak dilakukan pemangkasan (*pruning*). Oleh karena itu, pengaturan parameter seperti *max_depth*, *min_samples_split*, dan *criterion* menjadi krusial untuk menjaga keseimbangan antara akurasi dan generalisasi model.

2.4. Algoritma Random Forest

Random Forest adalah metode *ensemble learning* yang membangun sejumlah pohon keputusan untuk meningkatkan akurasi serta kestabilan prediksi. Setiap pohon dibentuk dari sampel data dan fitur yang dipilih secara acak, sehingga menghasilkan model yang lebih tahan terhadap *noise* dan mengurangi risiko *overfitting*. Hasil prediksi akhir diperoleh melalui mekanisme *majority voting* dari seluruh pohon yang terbentuk. Keunggulan utama algoritma ini terletak pada kemampuannya menangani dataset berukuran besar maupun yang tidak seimbang, serta menyediakan estimasi *feature importance* yang bermanfaat dalam analisis lebih lanjut.

Beberapa parameter yang berperan penting dalam performa Random Forest antara lain jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), serta jumlah fitur yang dipertimbangkan pada setiap pemisahan (*max_features*). Optimasi parameter-parameter tersebut diperlukan agar model mampu mencapai keseimbangan antara akurasi, generalisasi, dan efisiensi komputasi

2.5. Algoritma Grid Search

Grid Search merupakan teknik pencarian *hyperparameter* optimal dengan cara menguji seluruh kombinasi nilai yang telah ditentukan sebelumnya. Pendekatan ini termasuk dalam metode *exhaustive search*, di mana setiap kombinasi diuji secara sistematis dan dievaluasi menggunakan metrik performa, misalnya ROC-AUC atau PR-AUC.

Proses Grid Search biasanya dilakukan dengan membagi data ke dalam beberapa lipatan menggunakan *cross-validation*, sehingga model dilatih dan diuji secara bergantian untuk meminimalkan bias serta mengurangi risiko *overfitting*. Hasil akhir dari proses ini adalah konfigurasi parameter yang konsisten memberikan performa terbaik.

Dalam penelitian ini, parameter yang dioptimalkan meliputi:

- Decision Tree: *max_depth*, *min_samples_split*, *criterion*
- Random Forest: *n_estimators*, *max_depth*, *max_features*

Kelebihan utama Grid Search adalah kesederhanaan dan transparansi dalam proses

pencarian parameter. Namun, metode ini juga memiliki keterbatasan berupa kebutuhan komputasi yang tinggi jika ruang pencarian terlalu luas. Oleh karena itu, pemilihan ruang parameter yang efisien menjadi penting agar tuning tetap optimal tanpa memakan waktu berlebihan. Pada penelitian ini, Grid Search digunakan untuk memastikan setiap algoritma bekerja dengan konfigurasi terbaik, sehingga hasil evaluasi benar-benar mencerminkan performa maksimal dari masing-masing model klasifikasi.

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

Receiver Operating Characteristic (ROC) merupakan teknik visualisasi yang digunakan untuk menilai kinerja model klasifikasi biner. Kurva ROC menggambarkan hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai nilai ambang (threshold). TPR atau recall menunjukkan proporsi data positif yang berhasil dikenali model, sedangkan FPR merepresentasikan proporsi data negatif yang keliru diklasifikasikan sebagai positif. Kurva ROC digambarkan pada bidang dua dimensi, dengan sumbu X mewakili FPR dan sumbu Y mewakili TPR. Semakin dekat kurva ke sudut kiri atas grafik, semakin baik kemampuan model dalam membedakan kelas positif dan negatif.

Untuk mempermudah interpretasi, digunakan metrik AUC (Area Under Curve), yaitu luas di bawah kurva ROC. Nilai AUC berada pada rentang 0 hingga 1, dengan makna sebagai berikut:

- AUC = 1.0 → model ideal, mampu memisahkan kelas positif dan negatif tanpa kesalahan.
- AUC = 0.5 → performa setara dengan tebakan acak.
- AUC < 0.5 → model berkinerja buruk, bahkan lebih rendah dibanding tebakan acak.

$$TPR = \frac{TP}{TP + FN} \; dan \; FPR = \frac{FP}{FP + TN} \qquad (1)$$

2.7. Precision-Recall (PR) dan Area Under Curve (PR AUC)

Precision-Recall (PR) merupakan metode evaluasi yang menitikberatkan pada kinerja model terhadap kelas positif. Kurva PR menggambarkan hubungan antara nilai *Precision* dan *Recall* pada berbagai ambang batas (*threshold*). *Precision* menunjukkan persentase prediksi positif yang benar, sedangkan *Recall* mengukur proporsi data positif yang berhasil diidentifikasi oleh model.

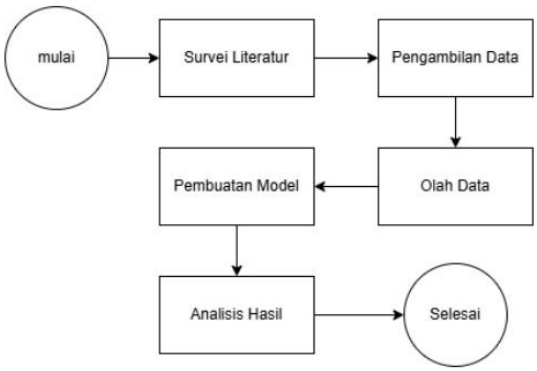
Metrik PR AUC atau luas di bawah kurva PR digunakan sebagai ukuran performa yang lebih peka terhadap distribusi data tidak seimbang. Dalam konteks *customer churn*, di mana jumlah nasabah yang berhenti jauh lebih sedikit dibandingkan yang bertahan, PR AUC memberikan gambaran yang lebih akurat mengenai kemampuan model dalam mengenali nasabah yang benar-benar berisiko churn.

Secara umum, rumus dasar yang digunakan dalam evaluasi PR adalah:

$$Precision = \frac{TP}{TP + FP} \; dan \; Recall = \frac{TP}{TP + FN} \qquad (2)$$

3. METODE PENELITIAN

Pelaksanaan penelitian ini dilakukan melalui serangkaian tahapan sistematis yang digambarkan pada Gambar 1. dalam bentuk diagram alir. Setiap tahap dirancang untuk mendukung proses pemodelan prediksi *customer churn* secara optimal dengan memanfaatkan algoritma Decision Tree dan Random Forest, yang selanjutnya disempurnakan melalui proses optimasi *hyperparameter* menggunakan teknik Grid Search.

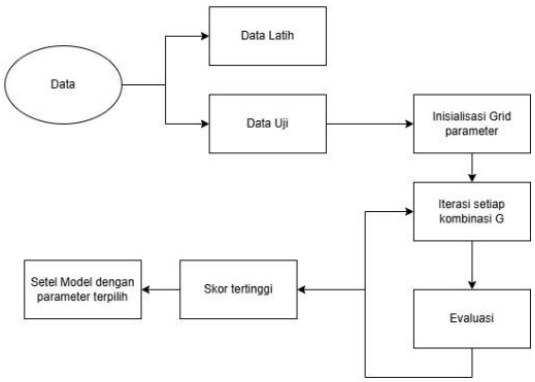


Gambar 1. Alur tahapan penelitian

3.1. Alur Grid Search

Grid Search dilakukan dengan melatih model pada setiap kombinasi parameter yang terdapat dalam ruang pencarian, kemudian mengevaluasi hasilnya menggunakan metrik yang telah ditentukan. Metrik evaluasi dapat berupa akurasi, presisi, *recall*, maupun ukuran lain yang relevan dengan tujuan pemodelan. Dari seluruh kombinasi yang diuji, metode ini akan memilih konfigurasi parameter yang memberikan performa terbaik sesuai indikator yang digunakan.

Tujuan utama penerapan Grid Search adalah menemukan set *hyperparameter* yang optimal sehingga model mampu mencapai kinerja maksimal secara keseluruhan. Dengan demikian, Grid Search membantu peneliti menentukan parameter yang paling sesuai dengan karakteristik data, sehingga model tidak hanya menghasilkan prediksi yang tinggi akurasinya, tetapi juga memiliki kemampuan generalisasi yang baik terhadap data baru.

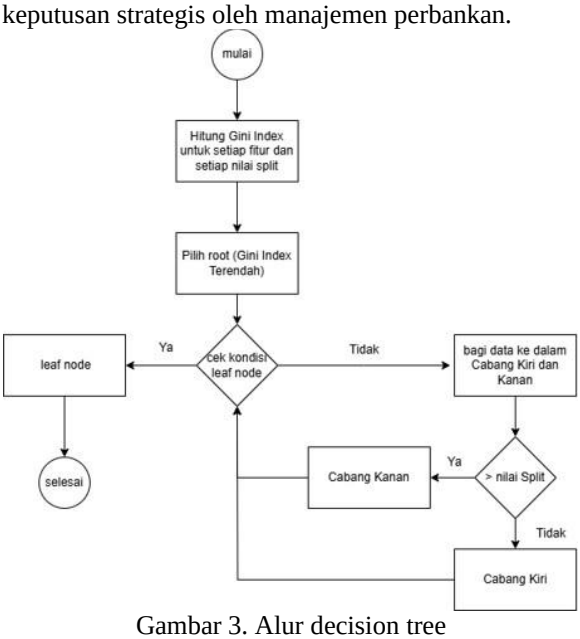


Gambar 2. Alur grid search

3.2. Alur Decision Tree

Algoritma Decision Tree bekerja dengan memecah data ke dalam cabang berdasarkan fitur yang paling relevan untuk menghasilkan keputusan yang terstruktur. Sementara itu, Random Forest sebagai metode *ensemble* menggabungkan banyak pohon keputusan sehingga mampu meningkatkan akurasi prediksi sekaligus mengurangi risiko *overfitting*.

Tahapan implementasi kedua model meliputi proses pra-pemrosesan data, pembagian data menjadi set pelatihan dan pengujian, pemilihan serta penyetelan *hyperparameter*, hingga tahap evaluasi dan perbandingan hasil prediksi. Seluruh rangkaian ini dirancang untuk memperoleh model yang optimal dalam memprediksi *customer churn* secara akurat, serta dapat digunakan sebagai dasar pengambilan keputusan strategis oleh manajemen perbankan.

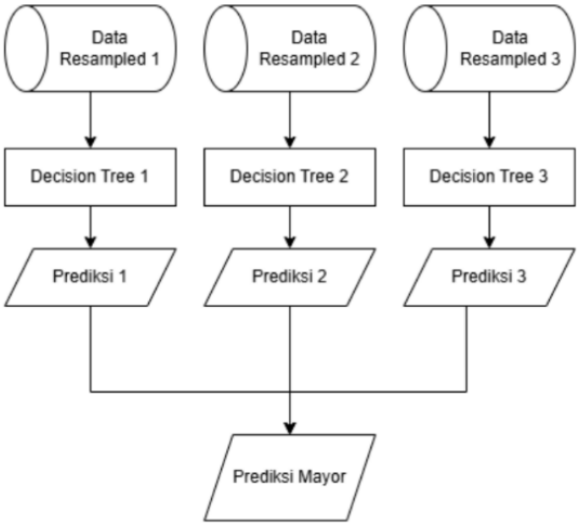


Gambar 3. Alur decision tree

3.3. Alur Random Forest

Random Forest merupakan algoritma ensemble yang membangun banyak pohon keputusan secara bersamaan untuk meningkatkan akurasi serta kestabilan prediksi. Model ini dikembangkan dari konsep Decision Tree, namun dengan pendekatan yang lebih kompleks karena setiap pohon dilatih menggunakan data hasil resampling melalui teknik Bootstrap Sampling.

Bootstrap Sampling sendiri adalah metode pengambilan sampel dengan pengembalian (replacement) dari dataset asli. Teknik ini memungkinkan terciptanya variasi pada data pelatihan tanpa harus menambah dataset baru. Dengan adanya variasi tersebut, Random Forest mampu menghasilkan model yang lebih stabil, memiliki akurasi lebih tinggi, serta lebih efektif dalam mengurangi risiko overfitting dibandingkan model yang hanya dilatih pada satu set data tunggal.



Gambar 4. Alur random forest

3.4. Input Data

Pada tahap awal penelitian, dilakukan proses *input* data dalam format CSV yang diunduh dari platform Kaggle. Dataset yang digunakan adalah Banking Customer Churn Prediction Dataset, yang terdiri atas 10.000 entri dengan 14 atribut. Rincian mengenai keempat belas kolom tersebut ditampilkan pada Tabel 1. sebagai deskripsi struktur data yang menjadi dasar analisis.

Tabel 1. Input data

No.	Nama Kolom	Keterangan	Jenis Data
1.	*RowNumber*	Nomor urut yang ditetapkan pada setiap baris dalam kumpulan data.	Integer
2.	*CustomerId*	Pengidentifikasi unik untuk setiap nasabah.	Integer
3.	*Surname*	Skor kredit nasabah.	String
4.	*CreditScore*	Pendapatan atau penghasilan pelanggan	Integer
5.	*Geography*	Lokasi geografis pelanggan (misalnya, negara atau wilayah).	String
6.	*Gender*	Jenis kelamin nasabah.	String
7.	*Age*	Usia nasabah.	Integer
8.	*Tenure*	Jumlah lama tahun nasabah telah menjadi nasabah bank.	Float
9.	*Balance*	Saldo akun nasabah.	Float
10.	*NumOfProducts*	Jumlah produk perbankan yang dimiliki nasabah.	Integer
11.	*HasCrCard*	Menunjukkan apakah pelanggan memiliki kartu kredit.	Biner
12.	*IsActiveMember*	Menunjukkan apakah nasabah adalah anggota aktif.	Biner
13.	*EstimatedSalary*	Perkiraan gaji nasabah.	Integer
14.	*Exited*	Menunjukkan apakah pelanggan telah keluar dari bank (biner: ya/tidak)	Biner

3.5. Skenario Pengujian

Penelitian ini berfokus pada analisis prediksi *customer churn* di sektor perbankan dengan memanfaatkan algoritma berbasis pohon keputusan sebagai pendekatan utama. Model prediksi dibangun menggunakan dua metode *ensemble learning*, yaitu Decision Tree dan Random Forest. Kedua algoritma tersebut dipilih karena masing-masing memiliki karakteristik serta keunggulan berbeda dalam mengolah data, sehingga mampu menghasilkan prediksi yang lebih akurat dan dapat diandalkan.

Tabel 2. Skenario pengujian

No.	Model Pohon Keputusan	Data Latih	Data Uji	Akurasi	ROC AUC	PR AUC
1.	Decision Tree	70	30	-	-	-
		80	20	-	-	-
2.	Random Forest	70	30	-	-	-
		80	20	-	-	-

4. HASIL DAN PEMBAHASAN

Setelah tahap pembangunan model selesai dijelaskan, penelitian dilanjutkan dengan analisis terhadap hasil pengujian yang dilakukan. Bagian ini menyajikan pembahasan terkait evaluasi performa masing-masing model berdasarkan uji coba yang telah dilaksanakan.

4.1. Hasil Pengujian Model

Pengujian terhadap dua model, yaitu Decision Tree dan Random Forest, dilakukan pada dua skenario pembagian data (70:30 dan 80:20) dan menghasilkan sejumlah temuan penting. Model Decision Tree menunjukkan akurasi konsisten sebesar 85% pada kedua skenario. Namun, nilai ROC-AUC yang diperoleh hanya 0.687 dan 0.619, serta PR-AUC sebesar 0.609 dan 0.612. Hal ini mengindikasikan bahwa meskipun akurasi cukup tinggi, kemampuan Decision Tree dalam membedakan kelas positif dan negatif masih terbatas dibandingkan model berbasis *ensemble*.

Sebaliknya, Random Forest mencatat akurasi sedikit lebih rendah, yakni 82% pada skenario 70:30 dan 81% pada skenario 80:20. Walaupun demikian, performa pada metrik ROC-AUC dan PR-AUC jauh lebih unggul dibandingkan Decision Tree, dengan nilai ROC-AUC berkisar 0.836–0.838 dan PR-AUC 0.647–0.650. Hasil ini menunjukkan bahwa Random Forest memiliki kemampuan generalisasi yang lebih baik, memberikan prediksi yang lebih stabil, serta lebih efektif dalam memisahkan kelas target, meskipun akurasi keseluruhan sedikit menurun.

Tabel 3. Hasil pengujian model

No.	Model + Parameter	Data Latih	Data Uji	Akurasi	ROC-AUC	PR-AUC
1.	Decision Tree (*max_depth=3, min_samples_split=2, min_samples_leaf=1)*	70	30	85%	0.687	0.609
		80	20	85%	0.619	0.612
2.	Random Forest (*n_trees=10, max_depth=5, max_bins=10, min_samples_split=2, min_samples_leaf=1)*	70	30	82%	0.836	0.647
		80	20	81%	0.838	0.65

4.2. Hasil Perbandingan Parameter

Berdasarkan hasil optimasi menggunakan Grid Search, konfigurasi terbaik untuk model Decision Tree diperoleh dengan parameter *max_depth = 5*, *min_samples_split = 2*, dan *min_samples_leaf = 1*. Hasil evaluasi dari parameter optimal tersebut ditunjukkan melalui *confusion matrix* yang dapat dilihat pada Gambar 5.

	precision	recall	f1-score	support
0	0.87	0.98	0.92	2416
1	0.80	0.40	0.53	584
accuracy			0.86	3000
macro avg	0.83	0.69	0.72	3000
ighted avg	0.86	0.86	0.84	3000

Gambar 5. Confusion matrix decision tree

Melalui proses optimasi menggunakan Grid Search, konfigurasi terbaik untuk model Random Forest diperoleh dengan parameter *n_trees = 7*, *max_depth = 7*, *min_samples_split = 1*, dan *min_samples_leaf = 5*. Evaluasi dari parameter optimal tersebut ditunjukkan melalui *confusion matrix* yang dapat dilihat pada Gambar 6.

	precision	recall	f1-score	support
0	0.88	0.97	0.92	2416
1	0.81	0.44	0.57	584
accuracy			0.87	3000
macro avg	0.84	0.71	0.75	3000
weighted avg	0.86	0.87	0.85	3000

Gambar 6. Confusion matrix random forest

5. KESIMPULAN DAN SARAN

Hasil pengujian menunjukkan bahwa penerapan Grid Search pada model Decision Tree maupun Random Forest mampu meningkatkan performa dalam memprediksi nasabah yang berisiko churn. Model Decision Tree mengalami peningkatan moderat dengan akurasi 0,86 dan *f1-score* kelas churn sebesar

0,53. Sementara itu, Random Forest memberikan hasil lebih baik dengan akurasi 0,87, *f1-score* kelas churn 0,57, serta nilai ROC-AUC tertinggi (0,838) dan PR-AUC (0,650). Temuan ini menegaskan bahwa optimasi *hyperparameter* pada algoritma *ensemble* seperti Random Forest lebih efektif dibandingkan Decision Tree dalam meningkatkan akurasi maupun kemampuan klasifikasi, sehingga Random Forest direkomendasikan sebagai pilihan utama untuk mendukung strategi retensi nasabah.

Untuk penelitian lanjutan, disarankan penggunaan metode optimasi yang lebih efisien, seperti Randomized Search, Bayesian Optimization, Genetic Algorithm, atau Optuna, guna mengatasi keterbatasan komputasi Grid Search. Selain itu, eksplorasi *feature engineering* melalui penambahan, transformasi, dan seleksi fitur dapat dilakukan untuk meningkatkan kualitas data.

DAFTAR PUSTAKA

[1] -, M. T. J., -, W., & -, F. D. M. (2025). Penerapan data mining algoritma decision tree untuk memprediksi keterlambatan pembayaran piutang usaha pada PT XYZ. *Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(3). https://doi.org/10.23960/jitet.v13i3.6828

[2] Abdurrahman, G., Oktavianto, H., & Sintawati, M. (2022a). Optimasi Algoritma XGBoost Classifier Menggunakan Hyperparameter Gridesearch dan Random Search Pada Klasifikasi Penyakit Diabetes. *INFORMAL: Informatics Journal*, *7*(3), 193. https://doi.org/10.19184/isj.v7i3.35441

[3] Abdurrahman, G., Oktavianto, H., & Sintawati, M. (2022b). Optimasi Algoritma XGBoost Classifier Menggunakan Hyperparameter Gridesearch dan Random Search Pada Klasifikasi Penyakit Diabetes. *INFORMAL: Informatics Journal*, *7*(3), 193. https://doi.org/10.19184/isj.v7i3.35441

[4] Alfianti, M. L. T., & Supriyanto, R. (2024). Perbandingan Kinerja Algoritma Random Forest, AdaBoost, dan XGBoost Dalam Memprediksi Resiko Penyakit Osteoporosis. *Jurnal Ilmu Komputer Dan Agri-Informatika*, *11*(2), 172–184. https://doi.org/10.29244/jika.11.2.172-184

[5] Anggraini, J. P., Chaya Gladys Zhafirah A, & Desiani, A. (2025). Perbandingan Algoritma Random Forest dan Extreme Gradient Boosting (XGBoost) dalam Klasifikasi Penyakit Gagal Jantung. *Komputika : Jurnal Sistem Komputer*, *14*(2), 149–157. https://doi.org/10.34010/komputika.v14i2.16618

[6] Ashraf, R. (2024). Bank customer churn prediction using machine learning framework. *Journal of Applied Finance & Banking*, 65–109. https://doi.org/10.47260/jafb/1445

[7] Bahtiar, A., & DP Silitonga, P. (2020). Penerapan algoritma Decision Tree untuk memprediksi penerima bantuan keluarga harapan. *Jurnal ICT : Information Communication & Technology*, *19*(1), 70–76. https://doi.org/10.36054/jict-ikmi.v19i1.93

[8] Dhika Malita Puspita Arum, Kartika Imam Santoso, Andri Triyono, Eko Supriyadi, Agus Susilo Nugroho, & Widodo, E. (2025). Algoritma Random Forest, Decision Tree, dan XGBoost untuk klasifikasi stunting pada balita. *Jurnal Transformatika*, *23*(1), 67–76. https://doi.org/10.26623/transformatika.v23i1.12202

[9] Fauzi, A., & Yunial, A. H. (2022). Optimasi Algoritma Klasifikasi Naive Bayes, Decision Tree, K – Nearest Neighbor, dan Random Forest menggunakan Algoritma Particle Swarm Optimization pada Diabetes Dataset. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *8*(3), 470. https://doi.org/10.26418/jp.v8i3.56656

[10] Hidayah, L., & Rosadi, M. I. (2024). Penerapan algoritma Random Forest untuk memprediksi jumlah santri baru. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3S1). https://doi.org/10.23960/jitet.v12i3s1.5237

[11] Karima, I. S. (2025). Penerapan Machine Learning untuk memprediksi Resiko Pengidap Penyakit Jantung menggunakan Algoritma decision tree. *Format : Jurnal Ilmiah Teknik Informatika*, *14*(1), 73. https://doi.org/10.22441/format.2025.v14.i1.007

[12] Louis Madaerdo Sotarjua, & Dian Budhi Santoso. (2022). Perbandingan algoritma knn, decision tree,*dan random*forest pada data imbalanced class untuk klasifikasi promosi karyawan. *Jurnal INSTEK (Informatika Sains Dan Teknologi)*, *7*(2), 192–200. https://doi.org/10.24252/instek.v7i2.31385