

Penerapan Algoritma Random Forest Untuk Klasifikasi Status Kredit Di Koperasi Prima Jaya Mandiri Bahorok

Primadana Fransiskus Ginting¹, Fati G. N. Larosa², Imelda Sri Dumayanti³

¹²³Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Info Artikel

Histori Artikel:

Received, Okt 27, 2025

Revised, Nov 11, 2025

Accepted, Nov 22, 2025

Keywords:

Random Forest, Klasifikasi, Status Kredit, Data Mining, Koperasi Prima Jaya Mandiri Bahorok

ABSTRAK

Penelitian ini bertujuan untuk menerapkan algoritma Random Forest dalam mengklasifikasikan status kredit anggota di Koperasi Prima Jaya Mandiri Bahorok guna mendukung proses pengambilan keputusan dalam penilaian kelayakan kredit. Permasalahan yang sering muncul dalam penentuan status kredit adalah kesalahan klasifikasi akibat banyaknya faktor yang saling berkaitan, seperti jumlah pinjaman, riwayat pembayaran, serta kondisi ekonomi anggota. Oleh karena itu, diperlukan metode klasifikasi yang memiliki kemampuan generalisasi tinggi dan akurasi yang andal. Metode yang digunakan dalam penelitian ini meliputi tahapan pengumpulan dan pra-pemrosesan data, penentuan label target (status lancar dan tidak lancar), pembagian data menjadi 70% data latih dan 30% data uji secara stratified, serta pelatihan model menggunakan algoritma Random Forest dengan konfigurasi terbaik, yaitu $n_estimators = 300$, $max_depth = 10$, dan $bootstrap$ aktif. Dari total 130 data kredit yang dianalisis, 91 data digunakan untuk pelatihan dan 39 data untuk pengujian. Hasil pengujian menunjukkan bahwa model Random Forest mampu mencapai tingkat akurasi sebesar 97,44% dan nilai F1-score sebesar 0,98 untuk kelas positif (status lancar). Berdasarkan confusion matrix dengan nilai $TN = 10$, $FP = 0$, $FN = 1$, dan $TP = 28$, model memiliki specificity sebesar 1,00 dan precision sempurna (1,00) pada kelas positif. Analisis feature importance menunjukkan bahwa variabel riwayat tunggakan bulanan merupakan faktor paling berpengaruh terhadap hasil klasifikasi, diikuti oleh jumlah pinjaman dan biaya hidup bulanan. Dengan demikian, penerapan algoritma Random Forest terbukti efektif dalam membantu proses klasifikasi status kredit di Koperasi Prima Jaya Mandiri Bahorok.

This is an open access article under the [CC BY-SA](#) license



Penulis Koresponden:

Primadana Fransiskus Ginting
Fakultas Ilmu Komputer,
Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: primadanagt@gmail.com

1. PENDAHULUAN

Koperasi memiliki peran yang sangat penting dalam memperkuat struktur perekonomian masyarakat Indonesia, khususnya pada sektor ekonomi kerakyatan di wilayah pedesaan dan semi-perkotaan. Berdasarkan data Kementerian Koperasi dan UKM tahun 2023, jumlah koperasi aktif di Indonesia tercatat lebih dari 127 ribu unit dengan total anggota mencapai lebih dari 27 juta orang.

Angka tersebut menunjukkan bahwa koperasi masih menjadi lembaga ekonomi yang relevan dan dipercaya oleh masyarakat. Di antara berbagai jenis koperasi yang ada, koperasi simpan pinjam menempati posisi yang cukup dominan karena berperan dalam menyediakan akses pembiayaan bagi masyarakat kecil, seperti petani, pedagang, serta pelaku usaha mikro yang sering kali belum memenuhi persyaratan lembaga perbankan formal. Melalui layanan yang relatif mudah, cepat, dan terjangkau, koperasi simpan pinjam berkontribusi dalam mendorong peningkatan kesejahteraan anggota.

Salah satu koperasi yang menjalankan fungsi tersebut adalah Koperasi Prima Jaya Mandiri Bahorok yang berlokasi di Kabupaten Langkat, Sumatera Utara. Mayoritas anggotanya berasal dari kalangan petani dan pelaku usaha skala kecil yang sangat bergantung pada fasilitas kredit untuk mendukung keberlangsungan usaha maupun memenuhi kebutuhan ekonomi sehari-hari. Meskipun demikian, dalam pelaksanaannya, proses analisis kelayakan kredit di koperasi ini masih menghadapi sejumlah kendala. Penilaian terhadap calon peminjam umumnya dilakukan berdasarkan pengalaman dan pertimbangan subjektif pengurus, tanpa didukung oleh sistem evaluasi berbasis data yang terstruktur. Kondisi tersebut berpotensi meningkatkan risiko kesalahan dalam pengambilan keputusan, terutama dalam menentukan apakah suatu pinjaman tergolong lancar atau berisiko bermasalah.

Secara umum, terdapat beberapa faktor yang dipertimbangkan dalam proses pengajuan kredit, seperti jenis pekerjaan, besaran pinjaman, jangka waktu angsuran, serta nilai jaminan yang diberikan[1]. Namun, tanpa adanya metode analisis yang sistematis, evaluasi terhadap variabel-variabel tersebut cenderung tidak konsisten dan rentan terhadap bias penilaian. Dampaknya, tingkat akurasi keputusan kredit menjadi kurang optimal dan berisiko menimbulkan kredit macet yang dapat mengganggu stabilitas keuangan koperasi serta menurunkan tingkat kepercayaan anggota[2].

Seiring dengan perkembangan teknologi informasi, penerapan metode berbasis machine learning menjadi salah satu solusi yang dapat digunakan untuk meningkatkan kualitas pengambilan keputusan. Salah satu algoritma yang memiliki kinerja baik dalam permasalahan klasifikasi adalah Random Forest. Algoritma ini merupakan metode ensemble learning yang membangun sejumlah pohon keputusan dan menggabungkan hasil prediksinya untuk menghasilkan model yang lebih stabil dan akurat[3], [4]. Keunggulan Random Forest terletak pada kemampuannya dalam menangani data dengan banyak variabel, mengurangi risiko overfitting, serta memberikan tingkat akurasi yang relatif tinggi dalam proses klasifikasi[5], [6].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma Random Forest dalam mengklasifikasikan status kredit anggota di Koperasi Prima Jaya Mandiri Bahorok ke dalam kategori Lancar dan Tidak Lancar. Diharapkan hasil penelitian ini dapat menjadi dasar dalam pengembangan sistem pendukung keputusan yang lebih objektif dan berbasis data, sehingga mampu meminimalkan risiko kredit bermasalah serta meningkatkan efektivitas manajemen kredit di lingkungan koperasi.

2. METODE PENELITIAN

Alur lengkap penelitian ini divisualisasikan pada Gambar 1 untuk memberikan gambaran yang lebih jelas mengenai urutan dan keterkaitan antar tahapan[7]. Berikut alur penelitian



Gambar 1. Framework Penelitian

Penelitian dimulai Peneliti melakukan tahap studi literatur. Tahap studi literatur merupakan kegiatan untuk mengumpulkan informasi dari berbagai sumber, dilanjutkan dengan pengumpulan data yang dilakukan di Koperasi Prima Jaya Mandiri Bahorok. Data yang digunakan adalah data Peminjam pada Januari 2024 – Agustus 2025. Kemudian dilakukan analisis menggunakan algoritma random forest. Setelah mendapatkan hasil maka akan dilakukan penentuan nilai akurasi dengan membandingkan data aktual dan data hasil klasifikasi hingga diperoleh hasil dan kesimpulan dari analisis yang dilakukan.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Metode Random Forest Secara Manual

Pada penelitian ini, untuk mempermudah pemahaman dalam proses pembentukan pohon keputusan random forest, dilakukan perhitungan manual dalam proses pembentukan pohon keputusan menggunakan metode CART (Classification and Regression Tree) dengan ukuran ketidakmurnian Gini Index[8].

Digunakan 20 data sample untuk perhitungan manual dengan kedalaman maksimum pohon ditetapkan sebesar 3 level, sehingga pola utama dari data dapat ditangkap dengan baik tanpa kehilangan kemampuan generalisasi. Untuk pohon pertama, atribut yang digunakan adalah Pekerjaan dan Status Pernikahan untuk memprediksi Status Kredit. Perhitungan Gini untuk setiap pekerjaan dapat dilihat pada persamaan berikut.

Pada evaluasi atribut Pekerjaan Suami, data dibagi menjadi empat kelompok: Karyawan (8 sampel), Pedagang (6), Petani (5), dan Wiraswasta (1). Untuk setiap kelompok, dihitung Gini node dari proporsi kelas Lancar (L) dan Tidak Lancar (TL)[9]. Hasil Perhitungan Gini atribut pekerjaan suami dapat dilihat pada Tabel 1

Tabel 1 Gini Atribut Pekerjaan Suami

No	Pekerjaan	Gini
1	Karyawan	$1 - (5/8)^2 - (3/8)^2 = 0.46875$
2	Pedagang	$1 - (5/6)^2 - (1/6)^2 = 5/18 = 0.27778$
3	Petani	$1 - (2/5)^2 - (3/5)^2 = 12/25 = 0.48$
4	Wiraswasta	$1 - (1)^2 - (0)^2 = 0$

$$\mathbf{Gini_{split}(Pekerjaan) = 0.390833}$$

Selanjutnya dilakukan pemisahan data berdasarkan atribut Status Pernikahan, dataset terbagi menjadi dua kelompok. Hasil perhitungan atribut status pernikahan dapat dilihat pada Tabel 2

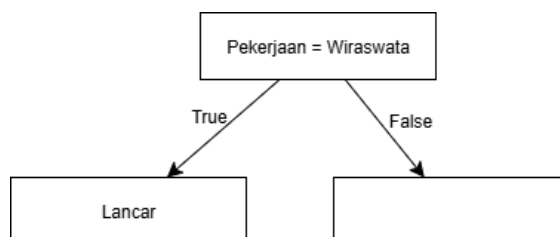
Tabel 2 Gini Atribut Status Pernikahan

No	Status	Gini
1	Lajang	$1 - (8/11)^2 - (3/11)^2 = 48/121 \approx 0.396694$
2	Menikah	$1 - (5/9)^2 - (4/9)^2 = 40/81 \approx 0.493827$

$$\mathbf{Gini_{split}(Status) = 0.440404}$$

Berdasarkan hasil perhitungan, nilai pembobotan Gini untuk atribut Pekerjaan adalah lebih kecil yaitu 0.390833 dibandingkan dengan atribut Status Pernikahan yaitu 0.440404. Karena prinsip pemilihan atribut pada algoritma CART adalah memilih atribut dengan nilai Gini Split terkecil (artinya tingkat ketidakmurnian paling rendah), maka atribut Pekerjaan dipilih sebagai pembuat node akar (root) pohon keputusan[10].

Selanjutnya dicek pada atribut pekerjaan gini yang paling rendah, diperoleh pada atribut karyawan = wiraswasta dengan hasil gini 0, dengan mayoritas peminjam lancar, dapat dilihat pada Gambar 2



Gambar 2 Pohon 1 (Level 1)

Setelah atribut Pekerjaan terpilih sebagai root, dilakukan pembuatan cabang dengan penjabaran berikut.

Evaluasi pekerjaan untuk level selanjutnya. Hasil Perhitungan Gini atribut pekerjaan suami level 2 dapat dilihat pada Tabel 3

Tabel 3 Gini Atribut Pekerjaan Suami Level 2

No	Pekerjaan	Gini
1	Karyawan	$1 - (5/8)^2 - (3/8)^2 = 0.46875$
2	Pedagang	$1 - (5/6)^2 - (1/6)^2 = 0.27778$
3	Petani	$1 - (2/5)^2 - (3/5)^2 = 0.48$
Gini_{split}(Pekerjaan) = 0.411404		

Selanjutnya dilakukan evaluasi status pernikahan untuk level selanjutnya. Hasil perhitungan atribut status pernikahan untuk level 2 dapat dilihat pada Tabel 4

Tabel 4 Gini Atribut Status Pernikahan Level 2

No	Status	Gini
1	Lajang	$1 - (8/11)^2 - (3/11)^2 = 0.396694$
2	Menikah	$1 - (4/8)^2 - (4/8)^2 = 0.5$
Gini_{split}(Status) = 0.466507		

Karena nilai Gini Split Pekerjaan lebih kecil, maka atribut Pekerjaan kembali dipilih sebagai pemisah terbaik pada percabangan berikutnya. Diperoleh pada atribut karyawan = pedagang dengan hasil gini 0.27778, dengan mayoritas peminjam lancar. Hasil Pohon setelah terbentuk level 1 dapat dilihat pada Gambar 3



Gambar 3 Pohon 1 (Level 2)

Selanjutnya melakukan evaluasi pekerjaan untuk level selanjutnya (Level 3). Hasil Perhitungan Gini atribut pekerjaan suami level 3 dapat dilihat pada Tabel 5

Tabel 5 Gini Atribut Pekerjaan Suami Level 3

No	Pekerjaan	Gini
1	Karyawan	$1 - (5/8)^2 - (3/8)^2 = 0.46875$

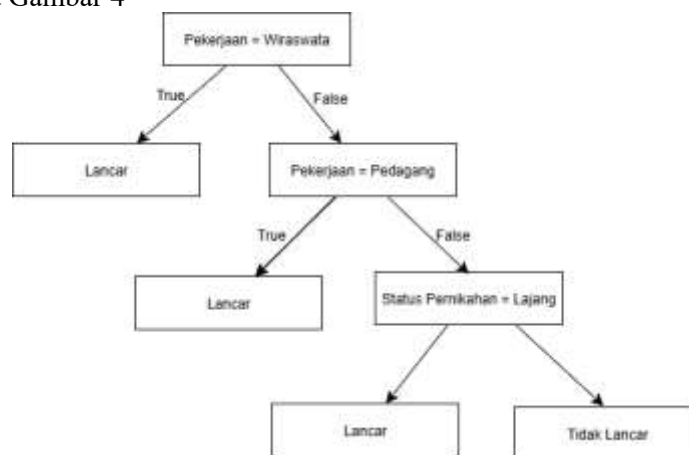
2	Petani	$1 - (2/5)^2 - (3/5)^2 = 0.48$
Gini_{split}(Pekerjaan) = 0.473077		

Selanjutnya dilakukan evaluasi status pernikahan untuk level selanjutnya. Hasil perhitungan atribut status pernikahan untuk level 3 dapat dilihat pada Tabel 6

Tabel 6 Gini Atribut Status Pernikahan Level 3

No	Status	Gini
1	Lajang	$1 - (5/7)^2 - (2/7)^2 = 0.408163$
2	Menikah	$1 - (4/6)^2 - (2/6)^2 = 0.444444$
Gini_{split}(Status) = 0.424908		

Karena nilai Gini Split Status Pernikahan lebih kecil, maka atribut Status Pernikahan dipilih sebagai pemisah terbaik pada percabangan berikutnya. Diperoleh pada atribut Status pernikahan = lajang dengan hasil gini 0.408163, dengan mayoritas peminjam lancar. Hasil Pohon setelah terbentuk level 3 dapat dilihat pada Gambar 4



Gambar 4 Pohon 1 (Level 3)

Karena pohon ditentukan dengan maksimum kedalam 3 level, sehingga proses perhitungan dan pembentukan pohon berakhir pada level 3.

Pembentukan pohon 2 dilakukan dengan cara yang sama seperti pada pohon 1. Atribut yang dievaluasi yaitu Jumlah Pinjaman dan Jangka Waktu. Split pada Jumlah Pinjaman (threshold 20.000.000). Hasil perhitungan atribut Split pada Jumlah Pinjaman (threshold 20.000.000) dapat dilihat pada Tabel 7

Tabel 7 Gini Atribut Jumlah Pinjaman (threshold 20.000.000)

No	Status	Gini
1	$\leq 20.000.000$	$1 - (11/13)^2 - (2/13)^2 = 0.260355$
2	$> 20.000.000$	$1 - (2/7)^2 - (5/7)^2 = 0.408163$
Gini_{split}(Jumlah Pinjaman) = 0.312088		

Split pada Jangka Waktu (threshold 15 bulan). Hasil perhitungan atribut Split pada Jangka Waktu (threshold 15 bulan) dapat dilihat pada Tabel 8

Tabel 8 Gini Atribut Jangka Waktu (threshold 15 bulan)

No	Status	Gini
1	≤ 15	$1 - (7/11)^2 - (4/11)^2 = 0.463$
2	> 15	$1 - (6/9)^2 - (3/9)^2 = 0.444444$
Gini_{split}(Jangka Waktu) = 0.454545		

Karena gini Jumlah Pinjaman lebih kecil maka dipilih sebagai root. Selanjutnya untuk level 2. Split lagi pada Jumlah Pinjaman (threshold 25.000.000). Hasil perhitungan atribut Split pada Jumlah Pinjaman (threshold 25.000.000) dapat dilihat pada Tabel 9

Tabel 9 Gini Atribut Jumlah Pinjaman (threshold 25.000.000)

No	Status	Gini
1	$\leq 25.000.000$	$1 - (0/2)^2 - (2/2)^2 = 0$
2	$> 25.000.000$	$1 - (2/5)^2 - (3/5)^2 = 0.48$
Gini_{split}(Jumlah Pinjaman) = 0.342857		

Selanjutnya Split pada Jangka Waktu (threshold 15). Hasil perhitungan atribut Split pada Jangka Waktu (threshold 15 bulan) dapat dilihat pada Tabel 10

Tabel 10 Gini Atribut Jangka Waktu (threshold 15 bulan)

No	Status	Gini
1	≤ 15	$1 - (1/3)^2 - (2/3)^2 = 0.444444$
2	> 15	$1 - (1/4)^2 - (3/4)^2 = 0.375$
Gini_{split}(Jangka Waktu) = 0.404762		

Untuk level 3, Split pada Jumlah Pinjaman (threshold 28.000.000). Hasil perhitungan atribut Split pada Jumlah Pinjaman (threshold 28.000.000) dapat dilihat pada Tabel 11

Tabel 11 Gini Atribut Jumlah Pinjaman (threshold 28.000.000)

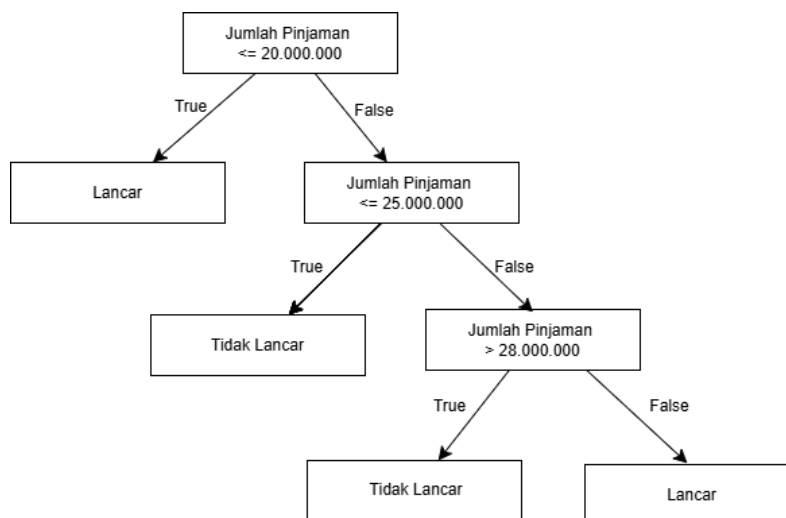
No	Status	Gini
1	$\leq 28.000.000$	$1 - (2/3)^2 - (1/3)^2 = 0.444444$
2	$> 28.000.000$	$1 - (0/2)^2 - (2/2)^2 = 0$
Gini_{split}(Jumlah Pinjaman) = 0.266667		

Split alternatif pada Jangka Waktu (threshold 15). Hasil perhitungan atribut Split pada Jangka Waktu (threshold 15 bulan) dapat dilihat pada Tabel 12

Tabel 12 Gini Atribut Jangka Waktu (threshold 15 bulan)

No	Status	Gini
1	≤ 15	$1 - (1/2)^2 - (1/2)^2 = 0.5$
2	> 15	$1 - (1/3)^2 - (2/3)^2 = 0.444444$
Gini_{split}(Jangka Waktu) = 0.466667		

Pada node Jumlah Pinjaman $> 25.000.000$, opsi pemisahan berikutnya dibandingkan kembali: (a) Jumlah Pinjaman ambang 28.000.000 vs (b) Jangka Waktu ambang 15 bulan. Memecah Jumlah Pinjaman menghasilkan kelompok $\leq 28.000.000$ (3 sampel, 2 Lancar & 1 Tidak Lancar, Gini = 0.444444) dan $> 28.000.000$ (2 sampel, 0 Lancar & 2 Tidak Lancar, Gini = 0), sehingga Gini berbobot = 0.266667. Adapun pemecahan Jangka Waktu pada node yang sama menghasilkan Gini berbobot = 0.466667. Karena $0.266667 < 0.466667$, dipilih kembali Jumlah Pinjaman (ambang 28.000.000). Dengan demikian, cabang $> 28.000.000$ menjadi node murni Tidak Lancar, sedangkan cabang $\leq 28.000.000$ terdominasi Lancar. Pohon 2 berdasarkan perhitungan diatas dapat dilihat pada Gambar 5



Gambar 5 Pohon 2

Secara keseluruhan, struktur Pohon 2 hingga level ke-3 adalah: akar Jumlah Pinjaman $\leq 20.000.000$ (didominasi Lancar), kemudian pada sisi $> 20.000.000$ dipisah $\leq 25.000.000$ (murni Tidak Lancar) vs $> 25.000.000$; dan pada $> 25.000.000$ dipisah $\leq 28.000.000$ (mayoritas Lancar) vs $> 28.000.000$ (murni Tidak Lancar).

Pembentukan pohon 3 dilakukan dengan cara yang sama seperti pada pohon 1 dan 2. Atribut yang dievaluasi yaitu Pendapatan Bulanan Suami dan Angsuran Bulanan.

Split pada Pendapatan Bulanan Suami (threshold 5.000.000). Hasil perhitungan atribut Pendapatan Bulanan Suami (threshold 5.000.000) dapat dilihat pada Tabel 13

Tabel 13 Gini Atribut Pendapatan Bulanan Suami (threshold 5.000.000)

No	Pendapatan	Gini
1	$\leq 5.000.000$	$1 - (7/10)^2 - (3/10)^2 = 0.42$
2	$> 5.000.000$	$1 - (6/10)^2 - (4/10)^2 = 0.48$
Gini_{split}(Pendapatan Suami) = 0.45		

Split pada Angsuran Bulanan (threshold 2.000.000). Hasil perhitungan atribut Pendapatan Angsuran Bulanan (threshold 2.000.000) dapat dilihat pada Tabel 14

Tabel 14 Gini Atribut Angsuran Bulanan (threshold 2.000.000)

No	Pendapatan	Gini
1	$\leq 2.000.000$	$1 - (8/12)^2 - (4/12)^2 = 0.444444$
2	$> 2.000.000$	$1 - (5/8)^2 - (3/8)^2 = 0.46875$
Gini_{split}(Angsuran Bulanan) = 0.454167		

Diperoleh: Nilai Gini Split untuk Pendapatan Bulanan Suami = 0.45

Nilai Gini Split untuk Angsuran Bulanan = 0.454167

Karena gini untuk Pendapatan Bulanan Suami lebih kecil maka dipilih sebagai root.

Selanjutnya untuk level 2 Split pada Pendapatan Bulanan Suami (threshold 7.000.000). Hasil perhitungan atribut Pendapatan Bulanan Suami (threshold 7.000.000) dapat dilihat pada Tabel 15

Tabel 15 Gini Atribut Pendapatan Bulanan Suami (threshold 7.000.000)

No	Pendapatan	Gini
1	$\leq 7.000.000$	$1 - (4/8)^2 - (4/8)^2 = 0.375$
2	$> 7.000.000$	$1 - (2/2)^2 - (0/2)^2 = 0$
Gini_{split}(Pendapatan Suami) = 0.3		

Split pada Angsuran Bulanan (threshold 2.000.000). Hasil perhitungan atribut Pendapatan Angsuran Bulanan (threshold 2.000.000) dapat dilihat pada Tabel 16

Tabel 16 Gini Atribut Angsuran Bulanan (threshold 2.000.000)

No	Pendapatan	Gini
1	$\leq 2.000.000$	$1 - (3/6)^2 - (3/6)^2 = 0.5$
2	$> 2.000.000$	$1 - (3/4)^2 - (1/4)^2 = 0.375$
Gini_{split}(Angsuran Bulanan) = 0.45		

Gini_{split}(Pendapatan Suami) = 0.3 < Gini_{split}(Angsuran Bulanan) = 0.45

Untuk level 3 Split pada Pendapatan Bulanan Suami (threshold 6.000.000). Hasil perhitungan atribut Pendapatan Bulanan Suami (threshold 6.000.000) dapat dilihat pada Tabel 17

Tabel 17 Gini Atribut Pendapatan Bulanan Suami (threshold 7.000.000)

No	Pendapatan	Gini
1	$\leq 6.000.000$	$1 - (3/4)^2 - (1/4)^2 = 0.375$
2	$> 6.000.000$	$1 - (1/4)^2 - (3/4)^2 = 0.375$
Gini_{split}(Pendapatan Suami) = 0.375		

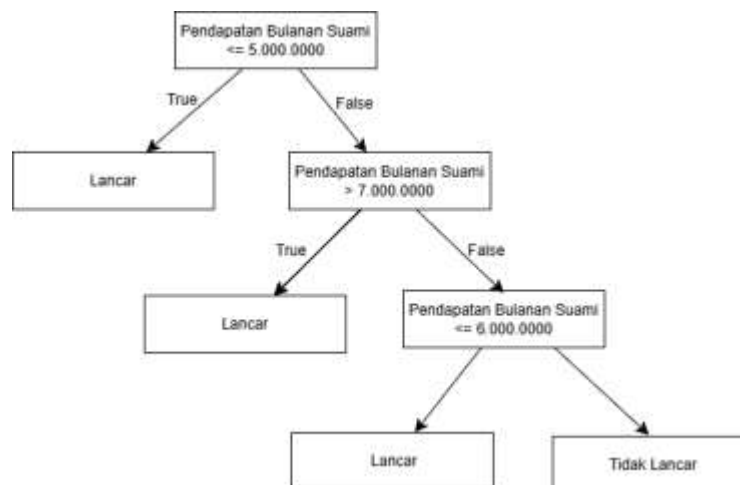
Split pada Angsuran Bulanan (threshold 2.000.000). Hasil perhitungan atribut Pendapatan Angsuran Bulanan (threshold 2.000.000) dapat dilihat pada Tabel 18

Tabel 18 Gini Atribut Angsuran Bulanan (threshold 2.000.000)

No	Pendapatan	Gini
1	$\leq 2.000.000$	$1 - (3/6)^2 - (3/6)^2 = 0.5$
2	$> 2.000.000$	$1 - (1/2)^2 - (1/2)^2 = 0.5$
Gini_{split}(Angsuran Bulanan) = 0.5		

Gini_{split}(Pendapatan Suami) = 0.375 < Gini_{split}(Angsuran Bulanan) = 0.5

Secara keseluruhan, struktur pohon keputusan yang terbentuk hingga level ke-3 menunjukkan bahwa atribut Pendapatan Bulanan Suami merupakan faktor dominan dalam menentukan status kelancaran kredit. Pohon 3 berdasarkan perhitungan diatas dapat dilihat pada Gambar 6



Gambar 6 Pohon 3

Akar pohon dimulai dari Pendapatan $\leq$ Rp5.000.000 (Lancar), kemudian pada cabang kanan Pendapatan $>$ Rp5.000.000 dilakukan pemisahan kembali pada Pendapatan $\leq$ Rp7.000.000 dan Pendapatan $>$ Rp7.000.000 (Lancar), serta pada bagian Pendapatan $\leq$ Rp7.000.000 dilakukan pemisahan lanjutan pada ambang Rp6.000.000 untuk membedakan antara nasabah Lancar dan Tidak Lancar. Proses pemilihan atribut pada setiap level selalu mengikuti prinsip dasar algoritma CART, yaitu memilih atribut dengan nilai Gini Split terkecil agar diperoleh pemisahan paling murni antara kelas Lancar dan Tidak Lancar[11].

Berdasarkan ketiga pohon yang sudah dibentuk diatas maka jika diterapkan dalam proses klasifikasi dengan contoh kasus: Responden nomor 20 berstatus menikah dengan pekerjaan suami sebagai Pedagang dan istri sebagai Karyawan. Suami memiliki pendapatan bulanan suami sebesar Rp5.500.000, sedangkan istri berpenghasilan Rp3.000.000, serta terdapat pendapatan tambahan lain sebesar Rp2.400.000. Total pendapatan keluarga digunakan untuk memenuhi biaya hidup bulanan sebesar Rp4.300.000. Responden memiliki jumlah pinjaman sebesar Rp28.000.000 dengan jangka waktu pengembalian selama 18 bulan, serta angsuran bulanan sebesar Rp3.666.000. Berdasarkan catatan kredit, responden tidak memiliki riwayat tunggakan dan menggunakan Sertifikat Rumah sebagai agunan. Dengan demikian, status kredit responden dikategorikan sebagai Lancar. Maka selanjutnya akan dicek terhadap ketiga pohon yang telah terbentuk tersebut: Pada pohon 1 masuk kedalam kategori Lancar karena pekerjaan suami = Pedagang.

Pada pohon 2 masuk kedalam kategori Lancar karena Jumlah pinjaman Rp28.000.000 (Tidak $>$ 28.000.000)

Pada pohon 3 masuk kedalam kategori Lancar karena Pendapatan Bulanan Suami = Rp5.500.000 ($\leq$ 6.000.000) Karena Kategori/Label Mayoritas adalah lancar maka nasabah tersebut masuk dalam kategori status kredit Lancar.

3.2 Implementasi Berbasis Python

Implementasi sistem dilakukan menggunakan bahasa pemrograman Python dengan framework Streamlit untuk membangun antarmuka pengguna berbasis web. Program dirancang untuk mengklasifikasikan Status Kredit (mis. Lancar/Tidak Lancar) pada Koperasi Prima Jaya Mandiri Bahorok menggunakan algoritma Random Forest.

3.2.1 Tampilan Awal

Aplikasi Random Forest Kredit Primadana dibangun dengan Streamlit sehingga seluruh fungsi utama tersaji dalam satu halaman (single-page) yang sederhana dan mudah dipahami. Pada bagian paling atas terdapat judul aplikasi disertai ikon, diikuti panel Petunjuk Singkat (dalam bentuk expander) yang merangkum alur penggunaan mulai dari unggah data hingga unduh lampiran hasil. Tampilan awal dapat dilihat pada Gambar 7

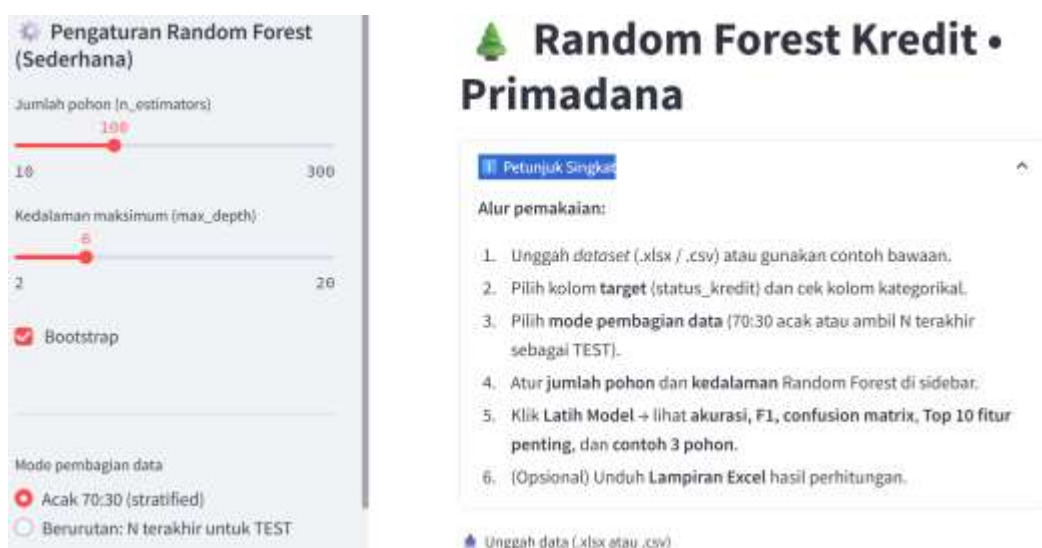


Gambar 7 Tampilan Awal

3.2.2 Tampilan Pengaturan Data dan Fitur

Pada sisi kiri layar tersedia pengaturan inti model yang mudah dipahami. Jumlah pohon ($n_estimators$) diatur lewat slider rentang 10–300 dengan nilai awal 100; parameter ini menentukan banyaknya pohon pada hutan—semakin besar umumnya varians model menurun (hasil lebih stabil) namun waktu pelatihan bertambah, sehingga disarankan mulai dari 100 dan naikan bertahap bila hasil masih fluktuatif. Kedalaman maksimum (max_depth) memiliki rentang 2–20 dengan nilai awal 6 untuk mengendalikan kompleksitas tiap pohon; semakin dalam, pola yang ditangkap makin detail tetapi risiko overfitting meningkat, sehingga dianjurkan memulai pada 6–10 dan menyesuaikan berdasarkan perbandingan kinerja TRAIN dan TEST. Tersedia pula opsi Bootstrap (kotak centang) yang, bila aktif, melatih setiap pohon pada sampel acak dengan pengembalian (bagging); pengaturan ini direkomendasikan karena membantu menurunkan varians dan meningkatkan kemampuan generalisasi model[12].

Di bagian bawah, pengguna memilih mode pembagian data. Opsi Acak 70:30 (stratified) membagi data secara acak menjadi 70% latih dan 30% uji sambil menjaga proporsi kelas—cocok untuk data yang tidak berurutan waktu. Alternatifnya, Berurutan: N terakhir untuk TEST mengambil N observasi paling akhir sebagai data uji dan sisanya sebagai data latih, cocok untuk data berurutan waktu (misalnya urutan pengajuan kredit per tanggal) agar evaluasi mencerminkan performa pada data terbaru nilai N dapat disetel di sidebar (bawaan sekitar 30).



Gambar 8 Tampilan Pengaturan Data

Pada penelitian ini pengaturan data dilakukan dengan penetapan jumlah pohon 300, Kedalaman 10, menggunakan bootstrap dan menggunakan mode pembagian data acak 70:30.

3.2.3 Tampilan Hasil Analisis

Setelah dilakukan pengaturan penetapan jumlah pohon 300, Kedalaman 10, menggunakan bootstrap dan menggunakan mode pembagian data acak 70:30. diperoleh hasil analisis dapat dilihat pada Gambar 9

Akurasi (TRAIN)	Akurasi (TEST)	F1 (TEST)	Jumlah Pohon
100.00%	97.44%	0.98	300
Konfusi TRAIN (0=Negatif, 1=Positif):		Konfusi TEST (0=Negatif, 1=Positif):	
	Pred 0	Pred 1	
Actual 0	23	0	
Actual 1	0	68	
	Pred 0	Pred 1	
Actual 0	10	0	
Actual 1	1	28	

Gambar 9 Hasil Analisis

Pembagian data latih dan data uji pada penelitian ini dilakukan dengan rasio sekitar 70% : 30%. Dari total 130 sampel data, sebanyak 91 sampel digunakan sebagai data latih dan 39 sampel digunakan sebagai data uji. Pada data latih, terdapat 23 sampel dengan kategori tidak lancar dan 68 sampel dengan kategori lancar. Sementara itu, pada data uji terdapat 10 sampel tidak lancar dan 29 sampel lancar.

Pada skenario split acak 70:30 (stratified) dengan pengaturan $n_estimators = 300$, $max_depth = 10$, dan bootstrap aktif, model mencapai akurasi TRAIN 100% dan akurasi TEST 97,44% dengan $F1 (TEST) = 0,98$ untuk kelas positif (sesuai label positif yang ditetapkan di aplikasi). Nilai ini menunjukkan keseimbangan precision–recall yang sangat baik pada data uji, dengan celah kinerja TRAIN–TEST sekitar 2,56% yang tergolong kecil sehingga model terbilang stabil.

Berdasarkan confusion matrix TEST, model menghasilkan $TN = 10$, $FP = 0$, $FN = 1$, $TP = 28$ (total 39 observasi). Artinya:

- Precision (positif) = $28 / (28+0) = 1,00 \rightarrow$ semua prediksi positif benar (tanpa FP).
- Recall (positif) = $28 / (28+1) \approx 0,97 \rightarrow$ hampir semua kasus positif berhasil dikenali, hanya ada 1 FN.
- Specificity (negatif) = $10 / (10+0) = 1,00 \rightarrow$ seluruh kasus negatif teridentifikasi dengan benar.

Pada TRAIN, matriks kebingungan menunjukkan klasifikasi sempurna ($TN=23$, $TP=68$, $FP=0$, $FN=0$), sehingga akurasi mencapai 100%. Hal ini wajar untuk Random Forest dengan jumlah pohon besar dan kedalaman yang cukup tinggi. Laporan Klasifikasi dan top 10 variabel yang paling berpengaruh juga di tampilkan pada tampilan website dapat dilihat pada Gambar 10

Classification Report (TEST)				
	precision	recall	f1-score	support
negatif	0.91	1.00	0.95	10
positif	1.00	0.97	0.98	29
accuracy			0.97	39
macro avg	0.95	0.98	0.97	39
weighted avg	0.98	0.97	0.97	39

Gambar 10 Tampilan Laporan Klasifikasi

Berdasarkan hasil pengujian dengan konfigurasi $n_estimators$ sebanyak 300, $max_depth = 10$, dan bootstrap aktif, model Random Forest menunjukkan kinerja yang sangat baik dengan akurasi uji sebesar 97,44%. Pada kelas negatif (tidak lancar), nilai precision mencapai 0,91 dengan recall 1,00 sehingga F1-score sebesar 0,95. Hal ini menunjukkan bahwa seluruh kasus negatif berhasil dikenali tanpa adanya kesalahan tipe false negative, meskipun terdapat sedikit penurunan pada precision yang mengindikasikan masih ada beberapa kasus positif yang diprediksi sebagai negatif. Sementara itu, pada kelas positif (lancar), nilai precision mencapai 1,00 dengan recall 0,97 sehingga menghasilkan F1-score sebesar 0,98. Artinya, semua prediksi positif yang dihasilkan model benar, meskipun

terdapat satu kasus positif yang tidak teridentifikasi dengan benar (false negative). Hasil analisis fitur menunjukkan bahwa faktor yang paling berpengaruh terhadap prediksi dapat dilihat pada Gambar 11

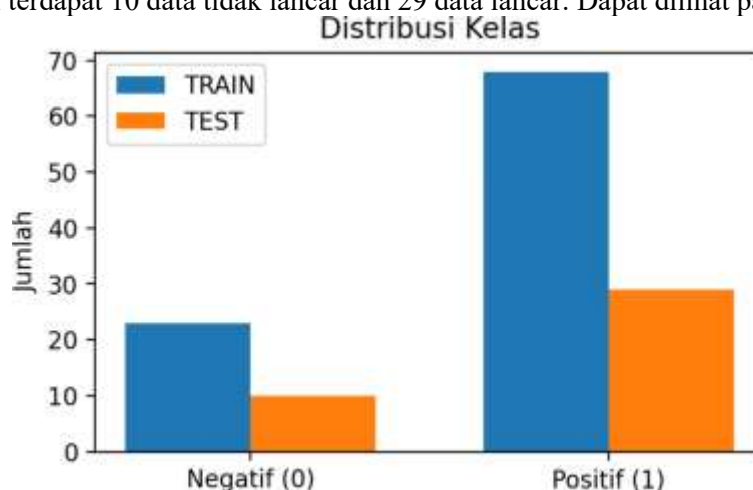
★ Top 10 Fitur Paling Berpengaruh

	feature	importance
25	riwayat_tunggakan_bulan	0.4983
21	biaya_hidup_bulanan	0.0681
22	jumlah_pinjaman	0.0595
18	pendapatan_bulanan_suami	0.0574
24	angsuran_bulanan	0.0573
19	pendapatan_bulanan_istri	0.0511
20	pendapatan_bulanan_lain	0.0475
23	jangka_waktu	0.0335
5	pekerjaan_suami_Petani	0.0116
4	pekerjaan_suami_Pedagang	0.0104

Gambar 11 Fitur Paling Berpengaruh

Faktor yang paling berpengaruh terhadap prediksi status kredit adalah riwayat_tunggakan_bulan dengan nilai kepentingan sebesar 0,4983, diikuti oleh biaya_bulanan (0,0681) dan jumlah_pinjaman (0,0595). Faktor-faktor lain seperti pendapatan, jangka waktu, serta jenis agunan juga turut memberikan kontribusi meskipun relatif lebih kecil. Hal ini selaras dengan logika bisnis, di mana riwayat tunggakan dan kemampuan bayar bulanan menjadi indikator utama kelancaran pembayaran kredit.

Hasil visualisasi distribusi kelas pada data latih (TRAIN) dan data uji (TEST) dalam proses klasifikasi status kredit menggunakan algoritma Random Forest. Terlihat bahwa dari total 130 sampel data, sebanyak 91 data digunakan untuk pelatihan (TRAIN) dan 39 data untuk pengujian (TEST) dengan proporsi pembagian 70:30 secara stratified. Pada data latih, terdapat sekitar 23 data dengan kelas negatif (status tidak lancar) dan 68 data dengan kelas positif (status lancar). Sementara itu, pada data uji terdapat 10 data tidak lancar dan 29 data lancar. Dapat dilihat pada Gambar 12



Gambar 12 Hasil Visualisasi Distribusi Kelas

Distribusi tersebut memperlihatkan bahwa jumlah data pada kedua kelas tidak seimbang, di mana kelas positif (status lancar) memiliki jumlah yang lebih besar dibandingkan kelas negatif (status tidak lancar). Meskipun demikian, penggunaan teknik stratified split memastikan bahwa proporsi kedua kelas tetap terjaga antara data latih dan data uji, sehingga model dapat belajar secara representatif terhadap pola dari masing-masing kategori. Visualisasi ini menegaskan bahwa pembagian data telah dilakukan dengan baik dan seimbang, sehingga mendukung keakuratan model dalam mengklasifikasikan status kredit anggota koperasi.

Menggunakan penetapan jumlah pohon 300, Kedalaman 10, menggunakan bootstrap dan menggunakan mode pembagian data acak 70:30 diperoleh sebanyak 3.428 rules yang terbetuk

berdasarkan pohon random forest data kredit nasabah koperasi koperasi prima jaya mandiri bahorok dapat dilihat pada Gambar 13

	tree_id	rule
3,419	300	Jika jumlah_pinjaman $\leq$ 44250000.000 DAN pekerjaan $\neq$ 'suami_Pedagang' DAN jangka_wa
3,420	300	Jika jumlah_pinjaman $\leq$ 44250000.000 DAN pekerjaan $\neq$ 'suami_Pedagang' DAN jangka_wa
3,421	300	Jika jumlah_pinjaman $\leq$ 44250000.000 DAN pekerjaan $\neq$ 'suami_Pedagang' DAN jangka_wa
3,422	300	Jika jumlah_pinjaman $\leq$ 44250000.000 DAN pekerjaan $\neq$ 'suami_Pedagang' DAN jangka_wa
3,423	300	Jika jumlah_pinjaman $\leq$ 44250000.000 DAN pekerjaan $\neq$ 'suami_Pedagang' DAN jangka_wa
3,424	300	Jika jumlah_pinjaman $\leq$ 44250000.000 DAN pekerjaan $\neq$ 'suami_Pedagang' $\rightarrow$ class = 0
3,425	300	Jika jumlah_pinjaman $>$ 44250000.000 DAN riwayat_tunggakan_bulan $\leq$ 2.500 $\rightarrow$ class = 1
3,426	300	Jika jumlah_pinjaman $>$ 44250000.000 DAN riwayat_tunggakan_bulan $>$ 2.500 DAN pendap
3,427	300	Jika jumlah_pinjaman $>$ 44250000.000 DAN riwayat_tunggakan_bulan $>$ 2.500 DAN pendap
3,428	300	Jika jumlah_pinjaman $>$ 44250000.000 DAN riwayat_tunggakan_bulan $>$ 2.500 DAN pendap

Gambar 13 Banyak Rules Yang Terbentuk

4. KESIMPULAN & SARAN

4.1 Kesimpulan

Berdasarkan hasil penelitian tentang penerapan algoritma Random Forest untuk klasifikasi status kredit di Koperasi Prima Jaya Mandiri Bahorok, dapat disimpulkan bahwa model yang dibangun mampu bekerja dengan sangat baik dan menghasilkan performa yang optimal. Proses penerapan algoritma dilakukan secara sistematis, dimulai dari tahap pengumpulan dan pra-pemrosesan data, penentuan label target (lancar dan tidak lancar), pembagian data menggunakan metode stratified dengan komposisi 70% data latih dan 30% data uji, hingga proses pelatihan model dengan parameter terbaik. Model Random Forest dengan konfigurasi $n_estimators = 300$, $max_depth = 10$, dan bootstrap aktif terbukti memberikan hasil klasifikasi yang paling akurat.

Dari total 130 sampel data, sebanyak 91 data digunakan sebagai data latih dan 39 data sebagai data uji. Hasil pengujian menunjukkan bahwa model memperoleh tingkat akurasi sebesar 97,44% dengan F1-score sebesar 0,98 untuk kelas positif (status lancar). Berdasarkan confusion matrix dengan nilai $TN = 10$, $FP = 0$, $FN = 1$, dan $TP = 28$, model mampu mengidentifikasi seluruh kasus negatif dengan benar (specificity = 1,00) serta mencapai precision sebesar 1,00 pada kelas positif, meskipun masih terdapat satu kasus positif yang tidak terklasifikasi secara tepat. Selain itu, hasil analisis feature importance menunjukkan bahwa riwayat tunggakan bulanan merupakan faktor yang paling berpengaruh dalam menentukan status kredit, diikuti oleh jumlah pinjaman dan biaya hidup bulanan.

4.2 Saran

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa saran untuk pengembangan penelitian selanjutnya agar hasil yang diperoleh menjadi lebih baik dan lebih komprehensif. Pertama, disarankan untuk menggunakan data historis dengan rentang waktu yang lebih panjang serta variasi kasus yang lebih beragam, seperti berbagai sektor pekerjaan dan jenis produk pinjaman, sehingga model yang dihasilkan menjadi lebih umum dan memiliki kemampuan generalisasi yang lebih baik. Penelitian selanjutnya dapat melakukan perbandingan dengan model lain, seperti Logistic Regression yang telah dikalibrasi maupun algoritma berbasis boosting seperti XGBoost atau LightGBM, guna memperoleh model dengan trade-off terbaik antara tingkat akurasi, stabilitas performa, dan kemudahan interpretasi hasil. Penting untuk melakukan standarisasi kualitas data melalui penyusunan data dictionary, validasi tipe data (angka dan teks), penanganan nilai yang hilang, serta audit terhadap outlier, khususnya pada data nominal rupiah, agar kualitas data yang digunakan benar-benar optimal dan dapat meningkatkan keandalan model klasifikasi.

REFERENSI

- [1] O.- Pahlevi, A.- Amrin, and Y.- Handrianto, "Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit," *J. Infortech*, vol. 5, no. 1, pp. 71–76, 2023, doi: 10.31294/infortech.v5i1.15829.
- [2] D. P. Siregar, J. Putri, U. Islam, and N. Sultanah, "Analisis strategi bank dalam menghadapi resiko kredit," vol. 5, no. 1, pp. 59–72, 2025.
- [3] S. D. Ahmad Fauzi, Nurlaelatul Maulidah, Riki Supriyadi, Hiya Nalatissifa, "Prediksi Harga Properti Di Indonesia Menggunakan Algoritma Random," vol. 4, no. 1, pp. 43–49, 2025.
- [4] T. Tambunan, M. Yohanna, and A. P. Silalahi, "Penerapan Metode Random Forest Dalam Mendeteksi Berita Hoax," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 7, no. 2, pp. 301–306, 2023, doi: 10.46880/jmika.vol7no2.pp301-306.
- [5] D. N. Handayani and S. Qutub, "Penerapan Random Forest Untuk Prediksi Dan Analisis Kemiskinan," vol. 4, no. 2, pp. 406–412, 2025.
- [6] A. P. Silalahi and H. G. Simanullang, *DECISSION TREE DAN PENERAPANNYA*. 2025.
- [7] D. Kiswanto, F. Ramadhani, N. M. Surbakti, and N. A. Nasution, "Pengembangan dan Implementasi Sistem Deteksi Serangan DDoS Berbasis Algoritma Random Forest," vol. 6, no. 3, 2025, doi: 10.47065/bit.v5i2.2203.
- [8] K. Neighbors, "Penerapan Data Mining Untuk Klasifikasi Penduduk Miskin Di Kabupaten Labuhanbatu Menggunakan Random Forest Dan," vol. 6, no. 2, pp. 23–36, 2025, doi: 10.47065/bit.v5i2.1783.
- [9] R. Forest, "Prediksi Kelayakan Pengajuan Pinjaman di Koperasi menggunakan," vol. 5, pp. 656–663, 2026.
- [10] H. Argretya, "Implementasi Random Oversampling pada Algoritma Random Forest untuk Klasifikasi Indeks Standar Udara di DKI Jakarta," vol. 10, no. 4, pp. 1189–1198, 2025, doi: 10.30591/jpit.v10i4.9466.
- [11] D. Hal *et al.*, "Penerapan Random Forest untuk Klasifikasi Risiko Penyakit Stroke Pada Rentang Usia 40-85 Tahun," vol. 8, pp. 118–124, 2025.
- [12] A. Rifai *et al.*, "OPTIMALISASI KLASIFIKASI INDEKS DESA MEMBANGUN MENGGUNAKAN METODE ENSEMBLE DAN ALGORITMA RANDOM FOREST," vol. 8, no. 4, pp. 8226–8234, 2024.