Terbit *online* pada laman : http://teknosi.fti.unand.ac.id/

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |

Artikel Penelitian

Perbandingan C4.5, *Artificial Neural Network*, dan Gaussian Naïve Bayes untuk Prediksi Piutang Pendidikan Tak Tertagih dengan Analisis *Grouped Permutation Importance*

Sry Faslia Hamka[a], *Amin Irmawan* [b]

[a,b]*Institut Teknologi dan Bisnis Muhammadiyah Wakatobi, Jln. Adhyaksa Nomor 37 Kecamatan Wangi-Wangi Selatan, Wakatobi 93791, Indonesia*

INFORMASI ARTIKEL

Sejarah Artikel:
Diterima Redaksi: 27 Juni 2026
Revisi Akhir: 20 Agustus 2026
Diterbitkan *Online*: 29 Agustus 2026

KATA KUNCI

piutang Pendidikan,

klasifikasi pembelajaran mesin,

C4.5,

artificial neural network,

grouped permutation importance

KORESPONDENSI

E-mail: sryfasliairmawan@gmail.com*

A B S T R A C T

Pengelolaan piutang biaya pendidikan merupakan tantangan operasional bagi perguruan tinggi, khususnya dalam mengidentifikasi secara dini mahasiswa yang berpotensi mengalami ketidaklancaran pembayaran. Penelitian ini bertujuan membandingkan kemampuan prediktif algoritma C4.5, *Artificial Neural Network* (ANN), dan Gaussian Naïve Bayes dalam mengklasifikasikan status piutang biaya pendidikan mahasiswa, serta menganalisis kontribusi atribut prediktor secara konsisten lintas-model. Data terdiri atas 258 rekaman mahasiswa dari Institut Teknologi dan Bisnis Muhammadiyah Wakatobi yang mengintegrasikan karakteristik akademik, wilayah tempat tinggal, kondisi sosial ekonomi wali, serta jumlah dan umur piutang UKT dan BPP. Evaluasi menggunakan holdout terstratifikasi 80:20 dan 90:10 serta *RepeatedStratifiedKFold* 5×10 sebagai prosedur utama. Seluruh algoritma dilatih pada pembagian data yang identik dengan tahap imputasi, encoding, standardisasi, dan *oversampling* ditempatkan dalam pipeline pelatihan. Kinerja dievaluasi menggunakan accuracy, precision, recall, F1-score, *balanced accuracy*, dan PR-AUC. Analisis interpretabilitas menggunakan grouped permutation importance diterapkan seragam pada ketiga algoritma. ANN memperoleh kinerja tertinggi dengan accuracy 82,09% ± 6,17%, recall 82,71%, F1-score 78,62%, *balanced accuracy* 82,20%, dan PR-AUC 91,32%. C4.5 menghasilkan kinerja mendekati ANN dengan *accuracy* 81,04% ± 4,98% dan PR-AUC 89,08%, disertai simpangan baku lebih kecil yang mencerminkan prediksi lebih stabil. Gaussian Naïve Bayes menghasilkan *recall* 80,58% namun *balanced accuracy* hanya 53,86% dan PR-AUC 46,99%, mengindikasikan kecenderungan prediksi positif berlebihan. Jumlah Piutang UKT Mahasiswa merupakan atribut paling berpengaruh secara konsisten pada ketiga algoritma, diikuti Umur Piutang UKT dan Jumlah Piutang BPP. Atribut akademik, wilayah, dan sosial ekonomi memberikan kontribusi yang kecil atau tidak stabil. ANN dipilih sebagai model utama untuk kepentingan operasional yang memprioritaskan deteksi piutang tak tertagih, sedangkan C4.5 menjadi alternatif kompetitif apabila institusi mengutamakan keterpahaman aturan klasifikasi. Hasil penelitian memberikan dasar ilmiah bagi pengembangan sistem pendukung keputusan prioritas pemeriksaan piutang dengan tetap mensyaratkan verifikasi manual sebelum tindak lanjut diambil.

1. PENDAHULUAN

Pembiayaan pendidikan tinggi merupakan salah satu komponen pengelolaan institusi yang memiliki dampak langsung terhadap keberlangsungan operasional dan kualitas layanan akademik. Di Indonesia, biaya pendidikan mahasiswa pada perguruan tinggi umumnya terdiri atas Uang Kuliah Tunggal (UKT) dan Biaya Penyelenggaraan Pendidikan (BPP) yang pemungutannya dilakukan secara periodik setiap semester. Ketidaklancaran pembayaran kedua komponen tersebut menimbulkan akumulasi piutang yang apabila tidak dikelola secara sistematis dapat memengaruhi arus kas institusi dan menghambat perencanaan anggaran. Pengelolaan piutang pendidikan yang efektif mensyaratkan kemampuan institusi untuk mengidentifikasi

https://doi.org/10.25077/TEKNOSI.v12i2.2026.312-326

secara dini rekaman mahasiswa yang berpotensi mengalami ketidaklancaran pembayaran, sehingga tindak lanjut penagihan dapat disusun berdasarkan prioritas yang terukur.

Pendekatan berbasis pembelajaran mesin telah banyak digunakan dalam berbagai konteks prediksi finansial, termasuk prediksi kelancaran pembayaran kredit [1], deteksi kualitas pengajuan kredit [2], dan prediksi gagal bayar pinjaman [3], [4]. Penelitian-penelitian tersebut menunjukkan bahwa algoritma klasifikasi mampu mengidentifikasi pola pada data historis dan menghasilkan prediksi yang dapat digunakan sebagai dasar pengambilan keputusan operasional. Di sisi lain, penerapan pembelajaran mesin pada konteks data keuangan pendidikan masih terbatas. Salah satu penelitian yang relevan adalah penerapan algoritma C4.5 untuk memprediksi keterlambatan pembayaran uang sekolah [5], yang menunjukkan potensi pohon keputusan dalam mengenali pola piutang pendidikan. Namun, penelitian tersebut hanya mengevaluasi satu jenis algoritma sehingga belum memberikan gambaran komparatif yang komprehensif antaralgoritma pada konteks yang sama.

Tiga algoritma yang umum dibandingkan dalam konteks klasifikasi finansial adalah C4.5 [6], [7] *Artificial Neural Network* (ANN) [8], [9] dan Gaussian Naïve Bayes [10]. C4.5 menghasilkan aturan klasifikasi berbentuk pohon keputusan yang dapat ditelusuri secara eksplisit [6], [7], ANN mampu memodelkan hubungan nonlinier yang kompleks antara prediktor dan kelas target [8], [9], sedangkan Gaussian Naïve Bayes menawarkan kesederhanaan komputasi dengan estimasi probabilitas berbasis Teorema Bayes [10]. Meskipun perbandingan ketiga algoritma ini telah dilakukan pada berbagai domain, seperti klasifikasi kelancaran pembayaran kredit [1] dan prediksi kelulusan mahasiswa [11], penerapannya secara terpadu pada data piutang pendidikan yang mengintegrasikan karakteristik akademik, kondisi sosial ekonomi wali, wilayah tempat tinggal, serta riwayat piutang UKT dan BPP belum banyak dieksplorasi.

Kesenjangan lain yang ditemukan pada penelitian terdahulu terletak pada aspek metodologis evaluasi dan interpretabilitas. Banyak penelitian melaporkan kinerja model berdasarkan satu pembagian data atau validasi silang tunggal tanpa memastikan bahwa seluruh algoritma menggunakan komposisi data latih dan data uji yang identik, sehingga perbedaan kinerja yang teramati sulit diatribusikan sepenuhnya pada karakteristik intrinsik algoritma [10]. Selain itu, perbandingan kepentingan atribut antara model yang berbeda sering kali didasarkan pada ukuran kepentingan internal masing-masing algoritma seperti *gain ratio* pada C4.5, bobot jaringan pada ANN, dan parameter distribusi pada Naïve Bayes yang tidak memiliki definisi maupun skala yang sebanding [12], [13]. Kondisi ini menyulitkan identifikasi atribut yang secara konsisten berpengaruh lintas-model. Di samping itu, sebagian besar penelitian menggunakan akurasi sebagai satu-satunya metrik evaluasi, padahal akurasi tidak sensitif terhadap ketidakseimbangan kelas dan tidak mencerminkan kemampuan model dalam mengenali kelas minoritas yang menjadi perhatian utama [14].

Penelitian ini bertujuan untuk mengatasi kesenjangan tersebut melalui tiga kontribusi utama. *Pertama*, membandingkan kemampuan prediktif algoritma C4.5, ANN, dan Gaussian Naïve Bayes dalam mengklasifikasikan status piutang biaya pendidikan mahasiswa menggunakan data dari Institut Teknologi dan Bisnis Muhammadiyah Wakatobi, dengan memastikan bahwa seluruh algoritma dilatih dan diuji pada komposisi data yang identik. *Kedua*, mengevaluasi kinerja model menggunakan metrik yang berorientasi pada kelas piutang tak tertagih, yaitu *recall*, F1-*score*, *balanced accuracy*, dan PR-AUC, sebagai pelengkap akurasi konvensional [14], [15]. *Ketiga*, menganalisis kontribusi setiap atribut prediktor menggunakan *grouped permutation importance* [12], [13] yang diterapkan secara konsisten dan seragam pada ketiga algoritma, sehingga perbandingan kepentingan atribut dapat dilakukan pada basis yang adil dan lintas-model.

Prosedur evaluasi utama menggunakan *RepeatedStratifiedKFold* 5×10 [10], [16], [17] yang menghasilkan 50 iterasi pengujian untuk setiap algoritma. Tahap imputasi, *encoding*, standardisasi, dan *oversampling* ditempatkan sepenuhnya dalam *pipeline* data latih [18], [19], [20] untuk mencegah kebocoran informasi dari data uji ke proses pelatihan. Hasil penelitian diharapkan dapat memberikan dasar ilmiah bagi pengembangan sistem pendukung keputusan yang membantu petugas keuangan perguruan tinggi menyusun prioritas pemeriksaan piutang secara lebih sistematis, terukur, dan dapat dipertanggungjawabkan.

2. METODE

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain observasional retrospektif dan eksperimen komparatif berbasis pembelajaran mesin untuk membandingkan kemampuan prediktif algoritma C4.5 [6], *Artificial Neural Network* (ANN) [9], dan Gaussian Naïve Bayes[10] dalam mengklasifikasikan status piutang biaya pendidikan mahasiswa ke dalam kelas tertagih dan tak tertagih. Ketiga algoritma dipilih karena ketiganya telah banyak dibandingkan pada konteks klasifikasi finansial [1], [2], sehingga relevan diterapkan pada konteks piutang pendidikan. Selain membandingkan kinerja prediktif, penelitian ini menganalisis kontribusi setiap atribut menggunakan *grouped permutation importance* [12], [13] yang diterapkan secara konsisten pada ketiga sehingga kepentingan atribut dapat dibandingkan menggunakan ukuran yang setara.

Unit analisis adalah satu rekaman mahasiswa yang mengintegrasikan karakteristik akademik, wilayah tempat tinggal, kondisi sosial ekonomi wali, serta jumlah dan umur piutang Uang Kuliah Tunggal (UKT) dan Biaya Penyelenggaraan Pendidikan (BPP). Seluruh rekaman yang memenuhi kriteria digunakan sebagai sampel (*total sampling*). Tahapan penelitian meliputi: pengumpulan dan penggabungan data; pemeriksaan konsistensi dan kelengkapan; penetapan dataset dan pengkodean kelas target; pemilihan atribut; pembentukan *pipeline* prapemrosesan; pelatihan dan evaluasi model C4.5, ANN, dan Gaussian Naïve Bayes menggunakan *holdout* 80:20, 90:10, dan *RepeatedStratifiedKFold* 5×10; serta analisis *grouped permutation importance* dan korelasi atribut keuangan utama.

2.2. Sumber, Periode, dan Pengumpulan Data

Data penelitian diperoleh dari dua sumber di Institut Teknologi dan Bisnis Muhammadiyah Wakatobi: Data akademik dari

Pangkalan Data Pendidikan Tinggi (PDDikti) dan sistem administrasi akademik perguruan tinggi, serta data keuangan dari Biro Administrasi Keuangan. Cakupan data mencakup tahun akademik 2020/2021, 2021/2022, dan 2022/2023 dan pencatatan piutang tahun 2020-2023. Status tertagih dan tak tertagih menggunakan status administratif resmi hasil rekonsiliasi keuangan institusi, dengan tanggal potong data pada Tahun Anggaran 2023. Ringkasan sumber dan cakupan data disajikan pada Tabel 1.

Tabel 1. Sumber dan cakupan data penelitian

Sumber Data	Informasi yang diperoleh	Periode
PDDikti dan sistem akademik perguruan tinggi	Identitas penghubung, program studi, dan karakteristik akademik mahasiswa	Tahun akademik 2020/2021–2022/2023
Data administrasi mahasiswa	Alamat kecamatan, alamat kelurahan/desa, pendidikan wali, pekerjaan wali, dan penghasilan wali	Tahun akademik 2020/2021–2022/2023
Biro Administrasi Keuangan	Jumlah piutang UKT, umur piutang UKT, jumlah piutang BPP, umur piutang BPP, dan status akhir piutang	Tahun 2020–2023

(Sumber: Diolah Penulis, 2026)

Penggabungan antarsumber dilakukan menggunakan NIM sebagai kunci pencocokan. NIM kemudian dikeluarkan dari dataset pemodelan karena tidak memiliki makna prediktif. Dari 260 rekaman awal, dua rekaman dikeluarkan karena tidak memiliki status target yang dapat ditetapkan, sehingga dataset akhir terdiri atas 258 rekaman. Rekaman diikutsertakan apabila tersedia di kedua sumber, dapat dicocokkan secara konsisten, berada dalam periode penelitian, atribut prediktornya tersedia atau dapat diimputasi, dan status akhir piutangnya telah ditetapkan oleh Biro Administrasi Keuangan.

2.3. *Definisi Kelas Target*

Variabel target adalah status akhir piutang yang telah ditetapkan oleh Biro Administrasi Keuangan, bukan ambang yang dibentuk ulang oleh peneliti. Status dikodekan sebagai variabel biner:

$$y_i = \begin{cases} 1, & \text{piutang tak tertagih} \\ 0, & \text{piutang tertagih} \end{cases}$$

Piutang tak tertagih ditetapkan sebagai kelas positif karena kelas tersebut menjadi fokus utama tindak lanjut penagihan. Dari 258 rekaman, 102 (39,53%) berstatus tak tertagih dan 156 (60,47%) berstatus tertagih. Hasil model ditafsirkan sebagai kemampuan klasifikasi berbasis pola historis, bukan sebagai bukti hubungan sebab-akibat antara prediktor dan status piutang [3], [4].

2.4. *Atribut Penelitian*

Sebanyak 10 atribut prediktor yang merepresentasikan karakteristik akademik, wilayah, sosial ekonomi wali, serta kondisi piutang UKT dan BPP disajikan pada Tabel 2. Umur piutang menggunakan kode administrative Biro Keuangan, yaitu kode 0 = lunas, 1 = jangka pendek, 2 = jangka menengah, dan 3 = jangka panjang. Ketidakseimbangan kelas ditangani melalui *RandomOverSampler* yang diterapkan terbatas pada data latih [18], [19], [20], [21].

Tabel 2. Definisi dan representasi atribut penelitian

No.	Atribut	Jenis Data	Representasi dalam pemodelan
1	Program Studi	Kategorikal nominal	*One-hot encoding*
2	Alamat Kecamatan	Kategorikal nominal	*One-hot encoding*
3	Alamat Kelurahan/Desa	Kategorikal nominal	*One-hot encoding*
4	Pendidikan Wali	Kategorikal	*One-hot encoding*
5	Pekerjaan Wali	Kategorikal nominal	*One-hot encoding*
6	Penghasilan Wali	Kategorikal/ordinal	*One-hot encoding*
7	Jumlah Piutang UKT Mahasiswa	Numerik, rupiah	Imputasi numerik dan standardisasi
8	Umur piutang UKT Mahasiswa	Ordinal	Pengkodean kategori dan standardisasi
9	Jumlah piutang BPP	Numerik, rupiah	Imputasi numerik dan standardisasi
10	Umur piutang BPP	Ordinal	Pengkodean kategori dan standardisasi

(Sumber: Diolah Penulis, 2026)

NIM, status mahasiswa, perguruan tinggi, jenjang, dan keterangan tidak digunakan sebagai prediktor karena tidak memiliki variasi nilai yang informatif atau hanya berfungsi sebagai pengenal.

2.5. *Pemeriksaan dan Penyiapan Data*

Pemeriksaan awal mencakup identifikasi duplikasi, ketidaksesuaian tipe data, perbedaan format penulisan, dan data hilang. Seluruh tahap imputasi, *encoding*, standardisasi, dan *oversampling* ditempatkan sepenuhnya di dalam *pipeline* pelatihan [10], [18], sehingga parameter transformasi hanya dipelajari dari data latih pada setiap pembagian atau *fold,* mencegah kebocoran informasi dari data uji [19], [20].

Nilai hilang pada atribut numerik dan ordinal diimputasi menggunakan nilai rata-rata data latih (*SimpleImputer*, *strategy="mean"*, lalu distandardisasi menggunakan

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (1)$$

dengan x_{ij} sebagai nilai atribut j pada rekaman i, sedangkan μ_j dan σ_j merupakan rata-rata dan simpangan baku atribut yang dihitung dari data latih [10]. Nilai hilang pada atribut kategorikal diimputasi menggunakan modus data latih kemudian dikonversi melalui *one-hot encoding* dengan handle_unknown="ignore" [10]. *RandomOverSampler* (sampling_strategy ="minority", random_state=0) diterapkan hanya pada data latih setelah pembagian, sehingga evaluasi selalu menggunakan distribusi kelas asli [18], [21]. Urutan *pipeline* pemodelan adalah:

Gambar 1. *Pipeline* pemodelan

Data uji hanya ditransformasi menggunakan parameter imputasi, *encoding*, dan standardisasi yang telah dipelajari dari data latih. Standardisasi tidak mengubah urutan pemisahan pada C4.5, tetapi tetap diterapkan agar seluruh algoritma menggunakan *pipeline* data yang konsisten [10], [18].

https://doi.org/10.25077/TEKNOSI.v12i2.2026.312-326

2.6. Pembentukan Model

Tiga algoritma dibangun menggunakan data latih dan data uji yang identik pada setiap scenario, dan diinisiasi ulang pada setiap pembagian *holdout* maupun *fold*.

2.6.1. Algoritma C4.5

C4.5 [6], [7] membentuk aturan klasifikasi hierarkis dengan memilih atribut berdasarkan *gain ratio* yang mengurangi bias terhadap atribut berkategori banyak [6]. Entropi, *information gain*, *split information,* dan *gain ratio* dihitung melalui:

$$Entropy\ (\mathrm{s}) = -\sum_{c=1}^{C} p_c log_2 p_c \quad (2)$$

$$Gain(S,A) = Entropy(S) - \sum_{v \in Value(A)} \frac{[S_v]}{[S]} Entropy(S_v) \quad (3)$$

$$SplitInformation(S,A) = -\sum_{v \in Value(A)} \frac{[S_v]}{[S]} log_2 \left(\frac{[S_v]}{S}\right) \quad (4)$$

$$GainRatio(S,A) = \frac{Gain(S,A)}{SplitInformation(S,A)} \quad (5)$$

2.6.2. Artificial Neural Network

ANN diimplementasikan menggunakan `MLPClassifier` [9], [10] dengan satu lapisan tersembunyi 100 neuron yang dilatih melalui *backpropagation* [8]. Nilai aktivasi neuron ke-j dihitung menggunakan:

$$h_j = f\left(\sum_{i=1}^{p} \omega_{ij} x_i + b_j\right) \quad (6)$$

Konfigurasi: aktivasi `relu`; *solver* `adam`, *learning rate* 0,001, *batch* otomatis; maksimum 200 iterasi, `alpha=0.0001`, `random_state=0` [9].

2.6.3. Gaussian Naive Bayes

Gaussian Naïve Bayes mengestimasi probabilitas kelas berdasarkan Teorema Bayes dengan asumsi independensi kondisional antarfitur [1], [2], [11].

$$P(c|X) \propto P(c) \prod_{j=1}^{p} P(x_j|c) \quad (7)$$

$$P(x_j|c) = \frac{1}{\sqrt{2\pi\sigma_{jc}^2}} \exp\left(-\frac{(x_j - \mu_{jc})^2}{2\sigma_{jc}^2}\right) \quad (8)$$

Diimplementasikan menggunakan `GaussianNB`, `var_smoothing=1×10⁻⁹`, prior diestimasi empiris dari data latih. Konfigurasi seluruh dirangkum pada Tabel 3.

Tabel 3. Konfigurasi model dan prosedur evaluasi

Komponen	Konfigurasi
C4.5	Kriteria pemisahan *gain ratio*; model dibangun kembali pada setiap *fold*
ANN	`MLPClassifier`; satu *hidden layer* berisi 100 neuron; aktivasi `relu`; *solver* adam; *learning rate* 0,001; *batch* otomatis; maksimum 200 iterasi; `alpha=0.0001; random_state=0`
Gaussian Naïve Bayes	`GaussianNB; var_smoothing=1e-9`; prior kelas diestimasi dari data latih
Imputasi numerik dan ordinal	Mean data latih
Imputasi kategorikal	Modus data latih
Encoding	*One-hot encoding*; kategori tak dikenal diabaikan
Standardisasi	`StandardScaler` berdasarkan data latih
Oversampling	`RandomOverSampler; sampling_strategy="minority"; random_state=0`
Ambang klasifikasi	0,50
Holdout	*Stratified* 80:20 dan 90:10
Validasi utama	*RepeatedStratifiedKFold 5 fold* × 10 pengulangan
Seed pembagian data	0
Pengulangan *permutation importance*	Lima pengacakan per atribut pada setiap *fold*
Metrik utama interpretabilitas	Penurunan PR-AUC
Metrik pendukung interpretabilitas	Penurunan *balanced accuracy*

(Sumber: Diolah Penulis, 2026)

2.7. Prosedur Pembagian Data dan Validasi

Evaluasi menggunakan dua skema *holdout* terstratifikasi (80:20 menghasilkan 206 data latih/ 52 data uji; 90:10 menghasilkan 232 data latih/ 26 data uji) dan *RepeatedStratifiedKFold* 5×10 sebagai prosedur utama [10], [16], [17]. Stratifikasi mempertahankan proporsi kelas pada setiap pembagian. Seluruh algoritma menggunakan indeks *fold* yang identik agar perbedaan kinerja hanya mencerminkan karakteristik algoritma.

RepeatedStratifiedKFold dikonfigurasi dengan `n_splits=5, n_repeats=10, random_state=0`, menghasilkan 50 evaluasi dengan data latih 206-207 rekaman dan data uji 51-52 rekaman per *fold*. Pada setiap *fold*: 1). *Pipeline* prapemrosesan dibangun ulang dari data latih; 2) *RandomOverSampler* diterapkan pada data latih; 3) model dilatih; 4) data uji ditransformasi menggunakan parameter dari data latih; dan 5) metrik dihitung pada distribusi kelas asli.

2.8. Metrik Evaluasi

Piutang tak tertagih ditetapkan sebagai kelas positif. Metrik yang digunakan meliputi *accuracy*, *precision*, *recall*, F1-*score*, *balanced accuracy*, dan PR-AUC [10], [14], [15]:

$$Accuracy = \frac{TP+T}{TP+TN+FP+FN} \quad (9)$$

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

$$F1 = 2\left(\frac{Precision \times Recall}{Precision+Reca}\right) \quad (12)$$

$$Specificity = \frac{TN}{TN+FP} \quad (13)$$

$$Balanced\ Accuracy = \frac{Recall+Specific}{2} \quad (14)$$

PR-AUC dihitung menggunakan `average_precision_score` dan lebih informatif dibandingkan ROC-AUC pada kondisi kelas tidak seimbang [10], [14]. Pada *holdout*, metrik dihitung satu kali, Pada *RepeatedStratifiedKFold*, dirata-ratakan dari 50 *fold*:

$$\bar{M} = \frac{1}{K}\sum_{k=1}^{K} M_k \quad (15)$$

$$SD = \sqrt{\frac{\sum_{k=1}^{K}(M_k - \bar{M})^2}{K-1}} \quad (16)$$

Proporsi kelas positif (39,53%) digunakan sebagai *baseline* deskriptif PR-AUC. Pemilihan model mempertimbangkan *recall*, F1-*score*, *balanced accuracy*, dan PR-AUC sebagai prioritas, tanpa pengujian signifikansi statistik antarmodel.

2.9. Analisis Interpretabilitas Model

Interpretabilitas ketiga algoritma dianalisis menggunakan *grouped permutation importance* yang bersifat *model-agnostic* [12], [13]. Metode ini dikembangkan dari konsep *permutation accuracy importance* yang pertama kali diperkenalkan oleh Breiman [12] dalam konteks *random forest*, kemudian diadaptasi sebagai metode interpretasi umum yang dapat diterapkan pada model klasifikasi apa pun, termasuk model dengan struktur internal berbeda seperti pohon keputusan, jaringan saraf, dan model probabilistik [13]. Metode yang sama diterapkan pada C4.5, ANN, dan Gaussian Naïve Bayes agar kepentingan atribut dapat dibandingkan menggunakan definisi dan metrik yang seragam.

Pengacakan dilakukan pada atribut asli sebelum imputasi, standardisasi, dan *one-hot encoding*. Dengan demikian, seluruh fitur hasil *one-hot encoding* yang berasal dari satu atribut, seperti kategori program studi atau alamat kelurahan/desa, diperlakukan sebagai satu kelompok atribut, sejalan dengan pendekatan *grouped permutation importance* untuk atribut kategorikal multi-level [13].

Analisis dilakukan pada setiap model dan setiap *fold RepeatedStratifiedKFold* melalui tahapan berikut: (1) menghitung PR-AUC dan *balanced accuracy* pada data uji asli sebagai nilai dasar; (2) memilih satu atribut asli; (3) mengacak urutan nilai atribut tersebut pada data uji; (4) mempertahankan seluruh atribut lain tanpa perubahan; (5) mentransformasi data yang telah diacak menggunakan *pipeline* dari data latih; (6) menghasilkan kembali probabilitas dan label prediksi tanpa melatih ulang model; (7) menghitung PR-AUC dan *balanced accuracy* setelah pengacakan; dan (8) mengulangi pengacakan sebanyak lima kali untuk setiap atribut [12].

Penurunan PR-AUC untuk atribut j, *fold* k, dan pengulangan $r$dihitung menggunakan:

$$PI_{jkr}^{PR} = PR - AUC_k^{baseline} - PR - AUC_{jkr}^{permuted} \quad (17)$$

$$PI_{jkr}^{BA} = BA_k^{baseline} - BA_{jkr}^{permuted} \quad (18)$$

Nilai positif menunjukkan bahwa pengacakan atribut menyebabkan penurunan kinerja sehingga model bergantung pada informasi atribut tersebut. Nilai mendekati nol menunjukkan bahwa atribut tidak memberikan kontribusi prediktif tambahan yang stabil. Nilai negatif menunjukkan bahwa kinerja pada pengacakan tertentu meningkat setelah informasi atribut dirusak [12].

Lima hasil pengacakan terlebih dahulu dirata-ratakan pada masing-masing *fold*:

$$\overline{PI}_{jk} = \frac{1}{5}\dot{\sum}_{r=1}^{5} PI_{jkr}$$

$$\overline{PI}_{j} = \frac{1}{50}\dot{\sum}_{r=1}^{50} \overline{PI}_{jk}$$

Simpangan baku dihitung berdasarkan 50 nilai pada tingkat *fold*. Persentase *fold* dengan penurunan PR-AUC dihitung menggunakan:

$$FoldPositif_j = \frac{\sum_{k=1}^{50} I(\overline{PI}_{jk}^{PR} > 0)}{50} \times 100\%$$

Penurunan PR-AUC digunakan sebagai ukuran kepentingan utama, sedangkan penurunan *balanced accuracy* digunakan sebagai ukuran pendukung. Atribut diperingkat berdasarkan rerata penurunan PR-AUC.

Seed pengacakan ditetapkan sebesar 0. Susunan indeks pengacakan yang sama digunakan pada ketiga algoritma untuk kombinasi *fold*, atribut, dan pengulangan yang sama.

Grouped permutation importance tidak ditafsirkan sebagai ukuran pengaruh kausal [12], [13]. Nilainya menunjukkan perubahan kemampuan prediksi ketika informasi pada suatu atribut dirusak dengan tetap mempertahankan atribut lainnya. Pada atribut yang saling berkorelasi kuat, nilai kepentingan dapat terbagi di antara beberapa atribut atau terkonsentrasi pada atribut tertentu yang lebih dominan digunakan oleh model [22].

2.10. Analisis Deskriptif dan Korelasi Atribut Utama

Analisis deskriptif dilakukan untuk membantu menjelaskan arah hubungan atribut yang memperoleh permutation importance tinggi. Distribusi atribut keuangan dibandingkan antara kelas piutang tertagih dan tak tertagih.

Median digunakan karena nilai nominal piutang berpotensi memiliki distribusi menceng dan mengandung nilai ekstrem. Median dihitung pada data asli sebelum standardisasi untuk Jumlah Piutang UKT Mahasiswa, Umur Piutang UKT Mahasiswa, Jumlah Piutang BPP, dan Umur Piutang BPP. Perbandingan median digunakan untuk mengidentifikasi kecenderungan arah hubungan prediktif dan tidak digunakan sebagai bukti hubungan sebab-akibat.

Hubungan antarprediktor keuangan dianalisis menggunakan korelasi peringkat Spearman [10]. Analisis tersebut digunakan karena umur piutang direpresentasikan dalam kategori ordinal dan hubungan antarprediktor tidak harus bersifat linear. Koefisien Spearman dihitung pada 258 rekaman data akhir sebelum *oversampling*. Pasangan atribut yang dianalisis meliputi Jumlah Piutang UKT Mahasiswa dengan Umur Piutang UKT Mahasiswa, serta Jumlah Piutang BPP dengan Umur Piutang BPP.

Koefisien Spearman dihitung menggunakan:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (19)$$

dengan $d_i$sebagai selisih peringkat dua variabel pada rekaman ke-i dan $n$sebagai jumlah rekaman.

 https://doi.org/10.25077/TEKNOSI.v12i2.2026.312-326

Apabila terdapat nilai yang sama, fungsi perangkat lunak memberikan peringkat rata-rata terhadap nilai yang berulang. Analisis korelasi digunakan untuk membantu menjelaskan kemungkinan pembagian atau pemusatan *permutation importance* pada atribut yang membawa informasi serupa.

2.11. Perangkat Lunak dan Pengendalian Reproduksibilitas

Pengolahan data, pemodelan, evaluasi, dan analisis interpretabilitas dilakukan menggunakan Python pada lingkungan Google Colaboratory.

Perangkat lunak dan pustaka utama yang digunakan meliputi: Python versi 3.10; pandas versi 1.5.3; NumPy versi 1.23.5; scikit-learn versi 1.2.2 [10]; imbalanced-learn versi 0.10.1 [18]; dan SciPy versi 1.11.3.

Reproduksibilitas dikendalikan melalui penggunaan *seed* sebesar 0 pada pembagian data, *RepeatedStratifiedKFold*, *RandomOverSampler*, ANN, dan *grouped permutation importance*. Indeks pembagian *holdout* dan seluruh 50 *fold* dibuat satu kali dan digunakan kembali untuk ketiga algoritma.

Objek *pipeline* dan model dibentuk kembali pada setiap *fold*. Dengan demikian, statistik imputasi, kategori *encoding*, parameter standardisasi, rekaman hasil *oversampling*, dan parameter model tidak terbawa dari *fold* sebelumnya.

2.12. Perlindungan Data dan Etika Penggunaan Model

Data mahasiswa dan wali mengandung informasi akademik, administratif, dan keuangan yang bersifat terbatas. NIM dan identitas langsung hanya digunakan pada tahap penggabungan data dan tidak dimasukkan ke dalam model.

Dataset pemodelan disimpan dalam bentuk terdeidentifikasi dan hanya dapat diakses oleh pihak yang memiliki kewenangan. Penggunaan data memperoleh izin dari Biro Administrasi Keuangan Institut Teknologi dan Bisnis Muhammadiyah Wakatobi.

Penelitian menggunakan data administratif sekunder yang bersifat retrospektif. Pengelolaan dan penggunaan data dilakukan sesuai dengan kewenangan institusi dan prinsip perlindungan kerahasiaan data.

Hasil prediksi dirancang sebagai informasi analitik pendukung bagi petugas keuangan dan bukan sebagai keputusan otomatis. Hasil model tidak digunakan secara langsung untuk menolak registrasi, membatasi layanan akademik, atau menjatuhkan sanksi kepada mahasiswa.

Setiap hasil prediksi harus diverifikasi melalui pemeriksaan bukti pembayaran, riwayat transaksi, komunikasi dengan mahasiswa, program keringanan yang berlaku, serta kondisi administratif aktual mahasiswa.

3. HASIL

3.1. Karakteristik Dataset

Dataset yang digunakan dalam pemodelan terdiri atas 258 rekaman mahasiswa yang diperoleh setelah proses pemeriksaan konsistensi dan penyiapan data. Berdasarkan distribusi kelas, sebanyak 102 rekaman (39,53%) tergolong dalam kategori piutang tak tertagih, sedangkan 156 rekaman sisanya (60,47%) tergolong dalam kategori piutang tertagih. Dalam proses evaluasi, piutang tak tertagih ditetapkan sebagai kelas positif (kode 1) karena kelas tersebut menjadi fokus utama dalam penentuan prioritas tindak lanjut penagihan.

Pemodelan menggunakan 10 atribut prediktor yang merepresentasikan karakteristik akademik mahasiswa, wilayah tempat tinggal, kondisi sosial ekonomi wali, serta karakteristik piutang UKT dan BPP. Nomor Induk Mahasiswa (NIM) tidak diikutsertakan sebagai prediktor karena berfungsi semata-mata sebagai pengenal unik mahasiswa. Atribut status mahasiswa, perguruan tinggi, jenjang, dan keterangan turut dikecualikan karena tidak memiliki variasi nilai yang secara informatif dapat mendukung proses klasifikasi.

Tiga algoritma yang dievaluasi, yaitu C4.5, *Artificial Neural Network* (ANN), dan Gaussian Naïve Bayes, diuji menggunakan dua skema pembagian data *holdout* terstratifikasi (80:20 dan 90:10) serta *RepeatedStratifiedKFold* (RSKF) lima lipatan dengan 10 pengulangan (5×10). Seluruh algoritma menggunakan pembagian data dan *fold* yang identik guna memastikan bahwa perbedaan kinerja yang teramati bersumber dari karakteristik intrinsik masing-masing model, bukan dari perbedaan komposisi data latih dan data uji. Proses imputasi, *encoding*, standardisasi, dan *oversampling* ditempatkan sepenuhnya dalam *pipeline* pelatihan sehingga parameter prapemrosesan hanya dipelajari dari data latih [18], [19], [20]. Kinerja seluruh algoritma pada setiap skenario pengujian dirangkum dalam Tabel 4.

Tabel 4. Kinerja model pada seluruh skenario pengujian

Model	Dataset	Train	Test	Kelas Positif	Acc.	Prec.	Recall	F1	Bal. Acc.	PR-AUC	Mean±SD
C4.5	Final 80:20	206	52	Tak tertagih (1)	76,92	69,57	76,19	72,73	76,80	86,13	-
ANN	Final 80:20	206	52	Tak tertagih (1)	76,92	65,52	90,48	76,00	79,11	84,46	-
Gaussian NB	Final 80:20	206	52	Tak tertagih (1)	48,08	42,11	76,19	54,24	52,61	42,37	-

C4.5	Final 90:10	232	26	Tak tertagih (1)	84,62	100,00	60,00	75,00	80,00	82,43	-	
ANN	Final 90:10	232	26	Tak tertagih (1)	80,77	77,78	70,00	73,68	78,75	85,29	-	
Gaussian NB	Final 90:10	232	26	Tak tertagih (1)	42,31	38,10	80,00	51,61	49,38	45,80	-	
C4.5	Final RSKF 5×10	206-207/*fold*	51-52/*fold*	Tak tertagih (1)	81,04	76,71	77,54	76,47	80,43	89,08	81,04 ± 4,98	
ANN	Final RSKF 5×10	206-207/*fold*	51-52/*fold*	Tak tertagih (1)	82,09	76,32	82,71	78,62	82,20	91,32	82,09 ± 6,17	
Gaussian NB	Final RSKF 5×10	206-207/*fold*	51-52/*fold*	Tak tertagih (1)	48,26	42,64	80,58	55,37	53,86	46,99	48,26 ± 8,90	

Catatan: Seluruh metrik dinyatakan dalam persen (%). Mean ± SD menunjukkan rata-rata akurasi beserta simpangan baku dari 50 *fold RepeatedStratifiedKFold*. Tanda "–" menunjukkan bahwa simpangan baku tidak dihitung pada pengujian *holdout* Tunggal.
(Sumber: Diolah Penulis, 2026)

3.2. Kinerja Model pada Pembagian Data 80:20

Pada skenario pembagian data 80:20, C4.5 dan ANN memperoleh akurasi yang setara, masing-masing sebesar 76,92%. Meski demikian, kedua model memperlihatkan karakteristik prediksi yang berbeda secara substansial. ANN mencatat *recall* sebesar 90,48%, melampaui C4.5 yang memperoleh 76,19%. Hasil ini mengindikasikan bahwa ANN mampu mendeteksi lebih banyak kasus piutang tak tertagih pada data uji.

Keunggulan ANN dalam mendeteksi kelas positif juga tercermin dari F1-*score* sebesar 76,00% dan *balanced accuracy* sebesar 79,11%, keduanya lebih tinggi dibandingkan C4.5 yang memperoleh F1-*score* 72,73% dan *balanced accuracy* 76,80%. Di sisi lain, C4.5 mencatat *precision* sebesar 69,57% dan PR-AUC sebesar 86,13%, sedikit lebih tinggi daripada ANN dengan *precision* 65,52% dan PR-AUC 84,46%. Dengan demikian, ANN bersifat lebih sensitif dalam mengidentifikasi kasus tak tertagih, sedangkan C4.5 lebih selektif dalam menetapkan suatu rekaman sebagai kelas positif.

Gaussian Naïve Bayes menghasilkan kinerja yang secara keseluruhan lebih rendah dibandingkan dua model lainnya, dengan akurasi 48,08%, *precision* 42,11%, F1-*score* 54,24%, dan *balanced accuracy* 52,61%. *Recall* sebesar 76,19% menunjukkan bahwa sebagian besar kasus tak tertagih masih dapat terdeteksi; namun rendahnya *precision* mengindikasikan tingginya laju *false positive*, yakni rekaman tertagih yang keliru diklasifikasikan sebagai tak tertagih.

3.3. Kinerja Model pada Pembagian Data 90:10

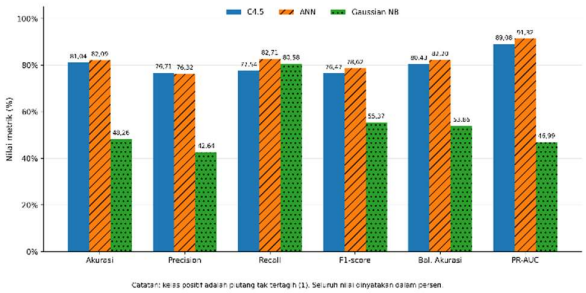
Pada skenario pembagian data 90:10, C4.5 memperoleh akurasi tertinggi sebesar 84,62% disertai *precision* sempurna sebesar 100,00%, yang berarti seluruh rekaman yang diprediksi sebagai piutang tak tertagih memang benar-benar termasuk dalam kelas tersebut. Namun, *recall* C4.5 hanya mencapai 60,00%, mengindikasikan bahwa model bersifat sangat selektif namun masih melewatkan sejumlah kasus piutang tak tertagih yang sesungguhnya.

ANN memperoleh akurasi 80,77%, *precision* 77,78%, *recall* 70,00%, F1-*score* 73,68%, dan *balanced accuracy* 78,75%. Meskipun akurasinya lebih rendah dari C4.5, ANN mencatat PR-AUC tertinggi pada skenario ini sebesar 85,29%, yang mengindikasikan kemampuan lebih baik dalam mempertahankan keseimbangan *precision* dan *recall* pada berbagai ambang klasifikasi.

Gaussian Naïve Bayes kembali menghasilkan *recall* yang relatif tinggi sebesar 80,00%, namun tidak disertai kinerja klasifikasi yang seimbang, sebagaimana tercermin dari akurasi 42,31%, *precision* 38,10%, F1-*score* 51,61%, dan *balanced accuracy* 49,38%. Pola ini menunjukkan kecenderungan model untuk memberikan prediksi positif secara berlebihan sehingga menghasilkan banyak kesalahan klasifikasi pada kelas tertagih.

3.4. Kinerja Model pada RepeatedStratifiedKFold

RepeatedStratifiedKFold (RSKF) 5×10 digunakan sebagai prosedur evaluasi utama karena setiap rekaman secara bergantian menjadi bagian dari data latih maupun data uji dalam 50 *fold* pengujian, sehingga menghasilkan estimasi kinerja yang lebih representatif dan stabil dibandingkan satu kali pembagian *holdout* [10], [16], [17]. Perbandingan visual kinerja ketiga model pada skenario ini disajikan pada Gambar 2.



Gambar 2. Perbandingan Kinerja Model pada *RepeatedStratifiedKFold* 5×10

Gambar 2 memperlihatkan bahwa ANN memperoleh nilai tertinggi pada sebagian besar metrik dalam evaluasi *RepeatedStratifiedKFold* 5×10, yaitu accuracy, recall, F1-score, *balanced accuracy*, dan PR-AUC. Meskipun demikian, perbedaan kinerja ANN dan C4.5 pada beberapa metrik relatif kecil. C4.5 menghasilkan precision sedikit lebih tinggi dan nilai kinerja yang mendekati ANN, sedangkan Gaussian Naïve Bayes menunjukkan recall yang relatif tinggi tetapi memiliki kinerja

https://doi.org/10.25077/TEKNOSI.v12i2.2026.312-326

keseluruhan yang lebih rendah. Oleh karena itu, hasil visual tersebut menunjukkan kecenderungan ANN memberikan keseimbangan kinerja yang lebih baik, tetapi belum membuktikan keunggulan yang signifikan secara statistik dibandingkan C4.5.

ANN memperoleh nilai rata-rata tertinggi pada sebagian besar metrik evaluasi dengan akurasi 82,09% ± 6,17%, *recall* 82,71%, F1-*score* 78,62%, *balanced accuracy* 82,20%, dan PR-AUC 91,32%. *Recall* yang relatif tinggi menunjukkan bahwa ANN mampu mengenali sebagian besar kasus piutang tak tertagih, sementara nilai *balanced accuracy* mengkonfirmasi bahwa kemampuan tersebut tidak diperoleh dengan mengorbankan kinerja pada kelas tertagih.

C4.5 menghasilkan kinerja yang mendekati ANN dengan akurasi 81,04% ± 4,98%, *precision* 76,71%, *recall* 77,54%, F1-*score* 76,47%, *balanced accuracy* 80,43%, dan PR-AUC 89,08%. Nilai *precision* C4.5 sedikit melampaui ANN, sedangkan simpangan baku akurasinya lebih kecil (4,98% berbanding 6,17%). Kombinasi kedua hal ini menunjukkan bahwa C4.5 menghasilkan prediksi yang lebih selektif dan relatif lebih stabil terhadap variasi komposisi *fold*.

Gaussian Naïve Bayes memperoleh akurasi 48,26% ± 8,90% dengan *recall* 80,58%, namun *precision* hanya 42,64%, F1-*score* 55,37%, *balanced accuracy* 53,86%, dan PR-AUC 46,99%. Kesenjangan besar antara *recall* dan *precision*, diperkuat oleh simpangan baku akurasi terbesar (8,90%), menunjukkan bahwa model cenderung memprediksi terlalu banyak kasus positif dan kinerjanya tidak stabil antara satu *fold* dan *fold* lainnya.

Berdasarkan hasil evaluasi RSKF, ANN memperoleh nilai rata-rata tertinggi pada accuracy, recall, F1-score, *balanced accuracy*, dan PR-AUC. Namun, selisih antara ANN dan C4.5 relatif terbatas, yaitu 1,05 poin pada accuracy, 1,77 poin pada *balanced accuracy*, dan 2,24 poin pada PR-AUC. Perbedaan yang lebih nyata terdapat pada recall, dengan ANN lebih tinggi 5,17 poin dibandingkan C4.5. Karena penelitian ini belum melakukan pengujian signifikansi statistik terhadap skor pada *fold* yang sama, hasil tersebut belum dapat digunakan untuk menyatakan bahwa ANN secara signifikan lebih unggul daripada C4.5. Dalam konteks penelitian ini, ANN dipilih sebagai model utama karena tujuan operasional lebih memprioritaskan kemampuan mendeteksi kasus piutang tak tertagih, sedangkan C4.5 tetap menjadi alternatif yang kompetitif ketika institusi lebih mengutamakan kestabilan model, precision, dan keterpahaman aturan klasifikasi.

3.5. Analisis Interpretabilitas dan Atribut yang Berpengaruh

Analisis interpretabilitas diterapkan secara konsisten pada ketiga algoritma, yaitu C4.5, *Artificial Neural Network* (ANN), dan Gaussian Naïve Bayes, menggunakan pendekatan *grouped permutation importance* [12], [13]. Pendekatan ini dilakukan dengan mengacak nilai satu atribut asli pada data uji sementara atribut lainnya dipertahankan tetap, kemudian mengukur perubahan kinerja model yang dihasilkan [12]. Pengacakan dilaksanakan sebelum tahap transformasi data sehingga seluruh fitur hasil *one-hot encoding* yang berasal dari atribut asal yang sama diperlakukan sebagai satu kelompok kesatuan, sejalan dengan pendekatan *grouped permutation importance* untuk atribut kategorikal multi-level [13]. Setiap atribut diacak sebanyak lima kali pada masing-masing dari 50 *fold RepeatedStratifiedKFold*.

Penurunan PR-AUC digunakan sebagai ukuran kepentingan utama, sejalan dengan fokus penelitian pada kemampuan model mengenali kelas piutang tak tertagih [14], [15]. *Balanced accuracy* digunakan sebagai ukuran pendukung untuk menilai dampak pengacakan atribut terhadap kemampuan model dalam mengenali kedua kelas secara seimbang. Nilai penurunan yang lebih besar mengindikasikan ketergantungan model yang lebih tinggi terhadap atribut tersebut, sebaliknya nilai yang mendekati nol atau negatif menunjukkan bahwa atribut tidak memberikan kontribusi prediktif tambahan yang stabil dalam model [12]. Atribut dengan *permutation importance* tertinggi pada setiap model dirangkum dalam Tabel 5.

Tabel 5. Atribut dengan *permutation importance* tertinggi pada setiap model

Model	Peringkat	Atribut	Penurunan PR-AUC, rerata ± SD (poin)	Penurunan *balanced accuracy*, rerata ± SD (poin)	*Fold* dengan penurunan PR-AUC
C4.5	1	Jumlah Piutang UKT Mahasiswa	46,85 ± 5,70	28,82 ± 7,25	100%
C4.5	2	Jumlah Piutang BPP	0,74 ± 1,65	1,12 ± 3,03	56%
C4.5	3	Alamat Kecamatan	0,18 ± 0,46	0,29 ± 0,82	40%
C4.5	4	Program Studi	0,08 ± 0,53	0,20 ± 1,11	40%
ANN	1	Jumlah Piutang UKT Mahasiswa	12,47 ± 6,12	6,21 ± 5,29	98%
ANN	2	Umur Piutang UKT Mahasiswa	9,86 ± 4,07	7,30 ± 4,30	100%
ANN	3	Jumlah Piutang BPP	2,79 ± 3,49	1,42 ± 3,48	84%
ANN	4	Umur Piutang BPP	1,06 ± 2,55	0,07 ± 3,18	68%
Gaussian Naïve Bayes	1	Jumlah Piutang UKT Mahasiswa	3,36 ± 3,79	1,91 ± 3,08	90%
Gaussian Naïve Bayes	2	Umur Piutang UKT Mahasiswa	1,49 ± 1,99	0,94 ± 1,49	78%
Gaussian Naïve Bayes	3	Alamat Kelurahan/Desa	0,85 ± 4,06	1,35 ± 5,79	58%
Gaussian Naïve Bayes	4	Jumlah Piutang BPP	0,62 ± 0,94	0,55 ± 1,25	74%

Catatan: Nilai menunjukkan penurunan kinerja setelah atribut diacak pada data uji, dinyatakan dalam poin persentase. Persentase *fold* menunjukkan proporsi dari 50 *fold* ketika pengacakan atribut menyebabkan penurunan PR-AUC.
(Sumber: Diolah Penulis, 2026)

Jumlah Piutang UKT Mahasiswa secara konsisten menempati peringkat pertama pada ketiga algoritma, sebagaimana diilustrasikan pada Gambar 2. Pada C4.5, pengacakan atribut tersebut menyebabkan PR-AUC menurun rata-rata sebesar 46,85 poin dan *balanced accuracy* menurun sebesar 28,82 poin, dengan penurunan yang terjadi pada seluruh *fold* (100%). Temuan ini mengindikasikan bahwa keputusan klasifikasi C4.5 sangat

bergantung pada informasi jumlah piutang UKT, sementara kontribusi atribut lain jauh lebih kecil, sehingga struktur pohon pada sebagian besar *fold* cenderung memusatkan pemisahan kelas pada atribut tersebut [6].

Pada ANN, kontribusi prediktif tersebar lebih merata pada beberapa atribut keuangan. Pengacakan Jumlah Piutang UKT Mahasiswa menurunkan PR-AUC sebesar 12,47 poin, sedangkan pengacakan Umur Piutang UKT Mahasiswa menurunkannya sebesar 9,86 poin; kedua atribut memberikan penurunan kinerja pada hampir seluruh *fold* (98% dan 100%). Jumlah Piutang BPP dan Umur Piutang BPP turut berkontribusi, meskipun dengan besaran yang lebih rendah dan variasi antar-*fold* yang lebih tinggi. Pola ini mengindikasikan bahwa ANN tidak hanya bergantung pada nilai nominal piutang, tetapi juga memanfaatkan informasi mengenai umur kewajiban pembayaran secara bersamaan [8], [9].

Pada Gaussian Naïve Bayes, Jumlah Piutang UKT Mahasiswa dan Umur Piutang UKT Mahasiswa juga menempati dua peringkat teratas, namun besaran penurunan PR-AUC setelah kedua atribut diacak jauh lebih kecil dibandingkan ANN maupun C4.5. Alamat Kelurahan/Desa menempati peringkat ketiga berdasarkan nilai rata-rata, tetapi simpangan bakunya yang jauh melampaui nilai rata-rata serta penurunan PR-AUC yang hanya terjadi pada 58% *fold* mengindikasikan bahwa kontribusi atribut ini tidak dapat dinyatakan stabil. Jumlah Piutang BPP memberikan pengaruh yang lebih kecil namun lebih konsisten dibandingkan atribut wilayah tersebut.

Secara keseluruhan, ketiga model menunjukkan kesepakatan bahwa Jumlah Piutang UKT Mahasiswa merupakan atribut prediktif paling penting. Umur Piutang UKT Mahasiswa dan Jumlah Piutang BPP membentuk kelompok atribut berikutnya yang memberikan kontribusi relatif konsisten, terutama pada ANN dan Gaussian Naïve Bayes. Sebaliknya, atribut program studi, alamat, pendidikan wali, pekerjaan wali, dan penghasilan wali menghasilkan penurunan kinerja yang kecil, tidak konsisten, atau mendekati nol setelah atribut keuangan dipertimbangkan oleh model [12], [13].

Hasil ini tidak dapat ditafsirkan sebagai indikasi bahwa karakteristik akademik, wilayah, dan sosial ekonomi tidak memiliki hubungan dengan kemampuan pembayaran mahasiswa. *Permutation importance* pada dasarnya mengukur kontribusi prediktif tambahan suatu atribut dengan mempertimbangkan keberadaan atribut lain dalam model. Dengan demikian, nilai yang rendah pada suatu atribut menunjukkan bahwa atribut tersebut tidak banyak menambah informasi prediktif baru setelah jumlah dan umur piutang telah dimasukkan ke dalam model.

Tabel 6. Median atribut keuangan menurut status piutang

Atribut	Piutang tak tertagih	Piutang tertagih
Jumlah Piutang UKT Mahasiswa	Rp. 3.000.000	Rp. 0
Umur Piutang UKT Mahasiswa	Kategori 3	Kategori 1
Jumlah Piutang BPP	Rp. 2.800.000	Rp. 2.000.000
Umur Piutang BPP	Kategori 3	Kategori 2

(Sumber: Diolah Penulis, 2026)

Arah hubungan atribut utama selanjutnya diperiksa melalui perbandingan distribusi menurut kelas target, sebagaimana dirangkum pada Tabel 6.

Rekaman piutang tak tertagih memiliki median jumlah dan umur piutang yang lebih tinggi dibandingkan rekaman piutang tertagih. Perbedaan paling besar tampak pada piutang UKT, dengan median sebesar Rp3.000.000 pada kelas tak tertagih berbanding Rp0 pada kelas tertagih. Median kategori umur piutang UKT juga berbeda secara nyata, yaitu kategori 3 pada kelas tak tertagih berbanding kategori 1 pada kelas tertagih. Pola ini mengindikasikan bahwa peningkatan nilai dan umur piutang berkaitan dengan meningkatnya kemungkinan suatu rekaman diklasifikasikan sebagai piutang tak tertagih. Perlu ditegaskan bahwa interpretasi tersebut merupakan hubungan prediktif (asosiatif) dan bukan menunjukkan hubungan sebab-akibat (kausal) [12], [13].

4. PEMBAHASAN

4.1. Interpretasi dan Perbandingan Kinerja Model

Hasil penelitian menunjukkan bahwa kinerja model dipengaruhi oleh skenario pembagian data yang digunakan. Pada skenario *holdout* 80:20, C4.5 dan ANN menghasilkan akurasi yang setara, sedangkan pada skenario 90:10 C4.5 mencatat akurasi lebih tinggi. Kendati demikian, hasil pada *holdout* 90:10 perlu ditafsirkan secara hati-hati mengingat data uji hanya terdiri atas 26 rekaman, di mana satu perubahan prediksi setara dengan pergeseran akurasi sekitar 3,85 poin persentase. Kondisi ini menyebabkan keunggulan berdasarkan satu pembagian *holdout* belum dapat dianggap mencerminkan kestabilan kinerja yang sesungguhnya [10].

Prosedur RSKF memberikan gambaran yang lebih konsisten mengenai kemampuan prediktif setiap model [10]. Penggunaan *fold* yang identik pada seluruh algoritma memastikan bahwa perbandingan dilakukan pada komposisi data latih dan data uji yang sama. Berdasarkan evaluasi ini, ANN memperoleh nilai rata-rata tertinggi pada *accuracy*, *recall*, F1-*score*, *balanced accuracy*, dan PR-AUC, sedangkan C4.5 menghasilkan *precision* sedikit lebih tinggi serta simpangan baku akurasi yang lebih kecil. Pola tersebut menunjukkan bahwa ANN cenderung lebih sensitif dalam mendeteksi piutang tak tertagih, sementara C4.5 cenderung lebih selektif dan stabil [6], [8]. Keunggulan nilai rata-rata ANN diduga berkaitan dengan kemampuannya merepresentasikan hubungan nonlinier antarprediktor [8], [9], tetapi penelitian ini belum menguji secara khusus bentuk hubungan tersebut maupun signifikansi perbedaan kinerja antarmodel.

Recall ANN sebesar 82,71% mengindikasikan bahwa sekitar delapan dari sepuluh kasus piutang tak tertagih berhasil dikenali, sebuah capaian yang relevan secara operasional mengingat kegagalan mendeteksi kasus tak tertagih (*false negative*) berpotensi menyebabkan institusi terlambat mengambil tindak lanjut penagihan [3], [4]. Namun demikian, *recall* tidak dapat dijadikan satu-satunya kriteria pemilihan model tanpa mempertimbangkan *precision*. Nilai *precision* ANN sebesar 76,32% menunjukkan bahwa sekitar tiga dari empat rekaman

https://doi.org/10.25077/TEKNOSI.v12i2.2026.312-326

yang ditandai sebagai tak tertagih memang benar termasuk kelas tersebut, sehingga tidak seluruh sumber daya verifikasi terbuang untuk kasus yang keliru. Keseimbangan keduanya tercermin dalam F1-*score* 78,62%, yang mengindikasikan bahwa ANN tidak mengorbankan salah satu metrik secara ekstrem demi memaksimalkan yang lain.

PR-AUC mengukur kemampuan model mempertahankan keseimbangan *precision* dan *recall* pada berbagai ambang klasifikasi, sehingga lebih informatif dibandingkan akurasi tunggal pada kondisi kelas tidak seimbang [14]. Relevansinya dalam konteks penelitian ini terletak pada kenyataan bahwa ambang klasifikasi 0,50 yang digunakan dalam pelatihan belum tentu optimal dalam penerapan operasional, dan kemampuan model pada berbagai ambang perlu dinilai. Menggunakan proporsi kelas positif sebesar 39,53% sebagai *baseline* acuan, ANN dan C4.5 masing-masing unggul sebesar 51,79 dan 49,55 poin persentase di atas *baseline* tersebut, mencerminkan kemampuan pemeringkatan kasus positif yang jauh melampaui klasifikasi acak. Sebaliknya, PR-AUC Gaussian Naïve Bayes hanya 7,46 poin di atas *baseline*, yang menegaskan bahwa *recall* tinggi model ini tidak mencerminkan kemampuan diskriminasi yang sesungguhnya, melainkan konsekuensi dari kecenderungan model menetapkan terlalu banyak rekaman sebagai kelas positif.

Perbedaan kinerja antara ANN dan C4.5 mencerminkan pertukaran (*trade-off*) mendasar antara sensitivitas dan selektivitas yang relevan untuk dipertimbangkan dalam konteks penerapan. C4.5 menghasilkan *precision* yang sedikit lebih tinggi (76,71% berbanding 76,32%) dan simpangan baku akurasi yang lebih kecil (4,98% berbanding 6,17%), yang mengindikasikan bahwa model ini lebih konsisten dalam menghasilkan prediksi positif yang tepat dan lebih stabil lintas komposisi *fold*. ANN, sebaliknya, unggul pada *recall* dengan selisih 5,17 poin, yang berarti lebih sedikit kasus tak tertagih yang terlewatkan. Pertukaran ini tidak menunjukkan bahwa salah satu model lebih baik secara absolut, melainkan bahwa pemilihan model sebaiknya disesuaikan dengan prioritas operasional institusi [6], [10].

Karakteristik ini memberikan implikasi praktis yang berbeda. ANN lebih sesuai apabila prioritas utama institusi adalah meminimalkan jumlah kasus tak tertagih yang tidak terdeteksi [3], [4]. Sebaliknya, C4.5 dapat dipertimbangkan apabila keterpahaman aturan klasifikasi oleh petugas keuangan menjadi pertimbangan penting, mengingat struktur pohon keputusan memungkinkan penelusuran pola atribut yang mendasari setiap keputusan klasifikasi [6], sedangkan keputusan ANN memerlukan metode interpretasi tambahan seperti SHAP atau *permutation importance* [12], [13].

Gaussian Naïve Bayes menunjukkan *recall* yang relatif tinggi, yaitu 80,58%, tetapi tidak diikuti oleh precision, F1-*score*, *balanced accuracy*, dan PR-AUC yang memadai. *Balanced accuracy* model ini hanya mencapai 53,86%, atau 3,86 poin persentase di atas tingkat acak sebesar 50% pada klasifikasi biner. Sebagai perbandingan, *balanced accuracy* ANN dan C4.5 masing-masing mencapai 82,20% dan 80,43%, yang menunjukkan kemampuan lebih seimbang dalam mengenali kelas tertagih maupun tak tertagih. Kesenjangan antara *recall* yang tinggi dan *balanced accuracy* yang mendekati tingkat acak mengindikasikan bahwa Gaussian Naïve Bayes terlalu sering menetapkan rekaman sebagai piutang tak tertagih. Dalam penerapan operasional, pola tersebut dapat menghasilkan daftar prioritas pemeriksaan yang terlalu luas dan meningkatkan beban verifikasi petugas keuangan [3], [4].

Temuan mengenai kecenderungan kinerja ANN yang lebih tinggi dibandingkan Gaussian Naïve Bayes sejalan dengan hasil penelitian Prayesy dan Negara [1] pada klasifikasi kelancaran pembayaran kredit, serta Nur dkk. [2] pada kualitas pengajuan kredit, yang keduanya menunjukkan bahwa jaringan saraf tiruan mampu menghasilkan kinerja lebih baik ketika hubungan antara prediktor dan kelas target bersifat kompleks dan non-linear. Adapun keunggulan C4.5 atas Naïve Bayes sejalan dengan temuan Rahmayanti dkk. [11] pada prediksi kelulusan mahasiswa. Persamaan kecenderungan ini tidak berimplikasi bahwa satu algoritma selalu lebih unggul secara universal, melainkan memperkuat kesimpulan bahwa kinerja model sangat dipengaruhi oleh karakteristik data, representasi atribut, distribusi kelas, dan prosedur validasi yang diterapkan [10].

Penggunaan *RandomOverSampler* dalam *pipeline* pelatihan berperan dalam mereduksi pengaruh ketidakseimbangan kelas pada data latih [18], [19], [20]. Penempatan *oversampling* setelah pembagian data memastikan bahwa duplikasi rekaman minoritas hanya berlaku pada data latih dan tidak mencemari data uji, yang merupakan prosedur penting untuk mencegah estimasi kinerja yang terlalu optimistis akibat kebocoran data [18], [19], [20], [21]. Selain itu, penggunaan *balanced accuracy* dan PR-AUC memberikan penilaian yang lebih menyeluruh dan tidak bias terhadap distribusi kelas dibandingkan akurasi konvensional [14], [15].

4.2. Perbandingan dengan Penelitian Terdahulu dan Kemajuan Penelitian

Konsistensi kecenderungan kinerja pada penelitian ini ANN dan C4.5 mengungguli Gaussian Naïve Bayes perlu ditempatkan dalam konteks literatur yang lebih luas untuk menilai apakah pola ini bersifat umum atau spesifik pada data penelitian ini. Prayesy dan Negara [1] serta Nur dkk. [2] melaporkan keunggulan jaringan saraf pada klasifikasi kredit, namun kedua penelitian tersebut menggunakan data debitur dengan atribut finansial yang lebih homogen dibandingkan dataset penelitian ini yang mengintegrasikan empat kelompok atribut berbeda. Kesamaan pola ini memperkuat dugaan bahwa jaringan saraf memiliki kapasitas generalisasi yang lebih baik pada data heterogen, meskipun pengujian hipotesis tersebut secara formal berada di luar cakupan penelitian ini. Demikian pula, keunggulan C4.5 atas Naïve Bayes yang sejalan dengan Rahmayanti dkk. [11] mengindikasikan bahwa pohon keputusan lebih mampu menangkap interaksi kondisional antaratribut yang relevan pada konteks data pendidikan, dibandingkan model yang mengasumsikan independensi antarprediktor.

Demikian pula, keunggulan C4.5 atas Gaussian Naïve Bayes memiliki kecenderungan yang sejalan dengan temuan Rahmayanti dkk. [11] pada prediksi kelulusan mahasiswa, yang mengindikasikan bahwa model pohon keputusan cenderung bekerja baik pada data yang mengandung hubungan bersyarat dan interaksi antarprediktor [6], [7]. Meski demikian, kesamaan kecenderungan antarpublikasi ini tidak dapat dijadikan dasar untuk menyimpulkan superioritas satu algoritma secara universal,

mengingat setiap penelitian memiliki jumlah data, distribusi kelas, representasi fitur, dan prosedur validasi yang berbeda-beda.

Prosedur evaluasi yang seragam memberikan dampak yang dapat diamati secara langsung pada hasil penelitian ini. Penggunaan *fold* yang identik memastikan bahwa perbedaan kinerja 1,05 poin akurasi dan 5,17 poin *recall* antara ANN dan C4.5 benar-benar mencerminkan perbedaan karakteristik algoritma, bukan artefak dari komposisi data yang berbeda. Penempatan *oversampling* di dalam *pipeline* pelatihan mencegah estimasi kinerja yang terlalu optimistis [18], [19], [20], [21] sebuah risiko yang nyata mengingat proporsi kelas minoritas hanya 39,53% dan duplikasi rekaman yang salah penempatan dapat secara artifisial meningkatkan metrik evaluasi. Dibandingkan dengan penelitian terdahulu yang umumnya menggunakan satu pembagian data atau validasi silang tunggal tanpa prosedur pencegahan kebocoran yang eksplisit, kerangka evaluasi penelitian ini menghasilkan estimasi kinerja yang lebih dapat dipercaya, meskipun dengan konsekuensi bahwa selisih kinerja yang teramati antara ANN dan C4.5 tergolong kecil dan belum dapat diklaim signifikan secara statistik [10].

Kontribusi lain penelitian ini terletak pada penggunaan metrik evaluasi yang berorientasi pada kelas piutang tak tertagih sebagai kelas minoritas yang menjadi perhatian utama. Pemilihan model tidak hanya didasarkan pada akurasi, melainkan juga mempertimbangkan *recall*, F1-*score*, *balanced accuracy*, dan PR-AUC secara komprehensif [14]. Pendekatan ini penting karena konsekuensi operasional akibat melewatkan kasus piutang tak tertagih berbeda secara substansial dari konsekuensi akibat menandai kasus tertagih sebagai tak tertagih [3], [4].

Sebagai konteks evaluasi, proporsi kelas positif dalam dataset tercatat sebesar 39,53%. Dibandingkan dengan *baseline* tersebut, PR-AUC ANN sebesar 91,32% dan C4.5 sebesar 89,08% mengindikasikan kemampuan pemeringkatan kasus positif yang kuat pada kedua model [14]. Sebaliknya, PR-AUC Gaussian Naïve Bayes sebesar 46,99% hanya sedikit melampaui proporsi kelas positif tersebut, sementara *balanced accuracy*-nya sebesar 53,86% juga hanya sedikit di atas tingkat acak (50%) [14], [15]. Temuan ini menegaskan bahwa *recall* Gaussian Naïve Bayes yang relatif tinggi tidak serta-merta mencerminkan kinerja klasifikasi yang baik, melainkan akibat dari kecenderungan model menghasilkan prediksi positif yang berlebihan dan banyak di antaranya keliru.

Secara keseluruhan, ANN memperoleh nilai rata-rata tertinggi pada akurasi, *recall*, F1-*score*, *balanced accuracy*, dan PR-AUC. Namun, selisih kinerja antara ANN dan C4.5 tergolong relatif kecil, yaitu 1,05 poin pada akurasi, 1,77 poin pada *balanced accuracy*, dan 2,24 poin pada PR-AUC. Keunggulan ANN yang paling menonjol terletak pada *recall*, dengan selisih 5,17 poin dibandingkan C4.5. Atas dasar ini, ANN dipilih sebagai model utama untuk kepentingan operasional yang memprioritaskan deteksi kasus tak tertagih [3], [4], sementara C4.5 tetap menjadi alternatif yang kompetitif ketika institusi lebih mengutamakan keterpahaman aturan klasifikasi dan kestabilan kinerja antar-*fold* [6].

Mengingat penelitian ini belum menyertakan pengujian signifikansi statistik terhadap skor kedua model pada *fold* yang sama, perbedaan kinerja antara ANN dan C4.5 belum dapat dinyatakan signifikan secara statistik [10]. Dengan demikian, simpulan bahwa ANN memiliki kinerja rata-rata tertinggi perlu dimaknai secara proporsional, dan tidak dimaksudkan untuk menyatakan superioritas ANN secara universal atas C4.5 pada seluruh konteks dataset piutang pendidikan.

4.3. Atribut yang Berpengaruh

Hasil analisis interpretabilitas mengindikasikan bahwa informasi yang berkaitan langsung dengan kondisi piutang memberikan kontribusi prediktif yang jauh lebih besar dibandingkan karakteristik akademik, wilayah, maupun sosial ekonomi [12], [13]. Jumlah Piutang UKT Mahasiswa konsisten menjadi atribut paling penting pada ketiga model, yaitu C4.5, ANN, dan Gaussian Naïve Bayes. Konsistensi lintas-algoritma ini memperkuat simpulan bahwa besaran kewajiban UKT merupakan indikator utama yang digunakan model dalam membedakan status piutang tertagih dan tak tertagih [5].

Meskipun ketiga model sepakat menempatkan Jumlah Piutang UKT Mahasiswa pada peringkat pertama, pola ketergantungan terhadap atribut tersebut berbeda secara karakteristik. C4.5 menunjukkan ketergantungan yang sangat terkonsentrasi, sebagaimana tercermin dari penurunan PR-AUC sebesar 46,85 poin setelah pengacakan atribut, yang mengindikasikan bahwa sebagian besar aturan pemisahan kelas pada struktur pohon bertumpu pada atribut ini [12]. Karakteristik tersebut sejalan dengan mekanisme intrinsik C4.5 yang memilih atribut dengan *gain ratio* tertinggi pada setiap simpul [6], [7]; ketika satu atribut memberikan pemisahan kelas yang sangat kuat, atribut lain yang membawa informasi serupa cenderung tidak lagi terpilih pada cabang pohon berikutnya.

ANN sebaliknya memperlihatkan pola interpretabilitas yang lebih tersebar. Selain Jumlah Piutang UKT Mahasiswa, Umur Piutang UKT Mahasiswa juga menghasilkan penurunan PR-AUC dan *balanced accuracy* yang besar dan konsisten antar-*fold*. Temuan ini mengindikasikan bahwa ANN memanfaatkan kombinasi nilai nominal dan durasi piutang secara simultan dalam membentuk representasi risiko [8], [9]. Jumlah dan umur piutang BPP turut memberikan kontribusi tambahan, meskipun dengan magnitudo yang lebih rendah. Pola ini mendukung dugaan bahwa hubungan antara kondisi piutang dan status ketertagihan tidak bergantung pada satu atribut tunggal, melainkan dapat melibatkan interaksi antarbeberapa indikator keuangan secara bersamaan [12].

Perbedaan pola kepentingan antara C4.5 dan ANN tersebut perlu dimaknai dengan mempertimbangkan struktur korelasi antarprediktor. Analisis korelasi Spearman mengungkapkan hubungan yang sangat kuat antara Jumlah Piutang UKT Mahasiswa dan Umur Piutang UKT Mahasiswa (koefisien = 0,969), serta hubungan kuat serupa antara Jumlah Piutang BPP dan Umur Piutang BPP (koefisien = 0,865). Pada kondisi atribut yang saling berkorelasi tinggi, nilai *permutation importance* berpotensi terbagi di antara beberapa atribut atau justru terkonsentrasi pada salah satu atribut yang lebih dominan dipilih model [22]. Karakteristik ini menjelaskan mengapa Umur Piutang UKT memberikan kontribusi besar pada ANN, namun

https://doi.org/10.25077/TEKNOSI.v12i2.2026.312-326

tidak menghasilkan penurunan kinerja berarti pada C4.5 setelah Jumlah Piutang UKT digunakan sebagai variabel pemisah utama [6], [12].

Gaussian Naïve Bayes juga menempatkan jumlah dan umur piutang UKT pada peringkat teratas, tetapi dengan kontribusi yang lebih kecil dan kurang stabil dibandingkan kedua model lainnya, sejalan dengan kinerja keseluruhan model yang relatif lebih rendah. Kondisi ini berkaitan dengan asumsi independensi kondisional pada Naïve Bayes yang kurang sesuai dengan korelasi sangat kuat antara jumlah dan umur piutang yang teramati pada dataset [22]. Ketika prediktor yang saling berhubungan diperlakukan seolah-olah independen, informasi yang substansinya sama dapat terhitung berulang kali atau justru tidak dimanfaatkan secara optimal oleh model. Penggunaan asumsi distribusi Gaussian pada fitur numerik, serta representasi biner hasil *one-hot encoding* pada fitur kategorikal, turut membatasi kesesuaian model terhadap karakteristik dataset penelitian ini [10].

Alamat Kelurahan/Desa menempati peringkat ketiga pada Gaussian Naïve Bayes, namun hasil tersebut tidak cukup stabil untuk dinyatakan sebagai atribut yang penting secara substantif. Nilai rerata penurunan PR-AUC hanya sebesar 0,85 poin dengan simpangan baku 4,06 poin, serta penurunan hanya teramati pada 58% *fold*. Variasi yang melebihi nilai rata-ratanya sendiri mengindikasikan bahwa peringkat atribut ini sangat dipengaruhi oleh komposisi data pada masing-masing *fold* [12], [13], sehingga atribut wilayah tidak dapat dijadikan dasar untuk menyusun simpulan substantif mengenai risiko piutang berdasarkan lokasi tempat tinggal.

Atribut pendidikan wali, pekerjaan wali, penghasilan wali, program studi, dan wilayah secara konsisten menghasilkan kontribusi tambahan yang kecil atau tidak stabil pada ketiga model. Temuan ini mengoreksi interpretasi awal yang sebelumnya cenderung menempatkan pendidikan dan pekerjaan wali sebagai atribut dominan berdasarkan ukuran kepentingan internal salah satu model. Ketika seluruh algoritma dianalisis menggunakan metode interpretabilitas yang seragam, dominasi tersebut tidak terkonfirmasi secara konsisten [12], [13]. Hasil ini menegaskan bahwa penilaian kepentingan atribut sebaiknya tidak didasarkan semata pada bobot internal satu model, terutama apabila jenis ukuran kepentingan yang digunakan berbeda-beda antarmodel dan tidak dapat diperbandingkan secara langsung.

Dibandingkan dengan pendekatan interpretasi pada penelitian-penelitian sebelumnya, kontribusi metodologis penelitian ini terletak pada penerapan *grouped permutation importance* secara seragam pada ketiga algoritma [12], [13]. *Feature importance* internal C4.5, bobot jaringan saraf pada ANN, dan parameter theta_ pada Gaussian Naïve Bayes pada dasarnya tidak memiliki definisi yang sebanding satu sama lain [6], [8], [10]. Penerapan metode interpretasi yang bersifat *model-agnostic* pada *fold* dan metrik evaluasi yang identik menghasilkan basis perbandingan yang lebih adil sekaligus memungkinkan identifikasi atribut yang konsisten pengaruhnya lintas-model [12], [13]. Pendekatan ini juga sejalan dengan kebutuhan interpretabilitas pada konteks prediksi gagal bayar, yang menekankan pentingnya penjelasan model sebelum hasil prediksi digunakan dalam proses operasional [3], [4].

Temuan mengenai dominasi atribut jumlah dan umur piutang memperluas cakupan penelitian terdahulu mengenai keterlambatan pembayaran pendidikan yang sebelumnya hanya menerapkan C4.5 pada satu jenis model [5]. Penelitian ini menunjukkan bahwa peran sentral atribut keuangan tidak hanya teramati pada model pohon keputusan, tetapi juga konsisten ditemukan pada ANN dan Gaussian Naïve Bayes ketika ketiganya dianalisis menggunakan prosedur interpretasi yang setara. Meskipun demikian, temuan tersebut tidak dapat ditafsirkan sebagai bukti bahwa jumlah atau umur piutang menyebabkan status tak tertagih secara kausal; nilai *permutation importance* hanya merefleksikan sejauh mana model memanfaatkan atribut tersebut dalam menghasilkan prediksi [12].

Secara operasional, hasil interpretabilitas ini mendukung penggunaan jumlah dan umur piutang UKT sebagai indikator utama dalam penyusunan daftar prioritas pemeriksaan piutang, dengan informasi piutang BPP berfungsi sebagai indikator pelengkap. Sebaliknya, atribut sosial ekonomi dan wilayah tidak sepatutnya digunakan secara langsung sebagai dasar penjatuhan sanksi atau pembatasan layanan terhadap mahasiswa, mengingat kontribusi prediktifnya yang tidak stabil dan berpotensi menimbulkan konsekuensi yang tidak adil bagi mahasiswa terkait [3], [4], [23], [24]. Hasil model tetap perlu diverifikasi melalui bukti pembayaran, riwayat komunikasi, program keringanan yang berlaku, serta kondisi aktual mahasiswa sebelum tindak lanjut diambil.

Interpretasi terhadap atribut jumlah dan umur piutang juga mensyaratkan bahwa definisi kelas target dibentuk secara independen dari kedua prediktor tersebut. Apabila status tak tertagih ditetapkan secara langsung menggunakan ambang jumlah atau umur piutang yang sama dengan yang digunakan sebagai prediktor, tingginya nilai *permutation importance* yang teramati dapat semata-mata mencerminkan aturan pembentukan label, bukan kemampuan model dalam memprediksi kondisi masa depan yang sesungguhnya [12], [13]. Oleh karena itu, waktu pengukuran prediktor dan waktu penetapan status akhir piutang perlu didokumentasikan secara eksplisit dan jelas guna memastikan tidak terjadi kebocoran target (*target leakage*) dalam proses pemodelan [18], [21].

4.4. Implikasi Penerapan

ANN dapat difungsikan sebagai komponen analitik pendukung pengambilan keputusan bagi petugas keuangan dalam menyusun daftar prioritas pemeriksaan piutang [3], [4], [23], [24]. Hasil prediksi tidak diperkenankan dijadikan dasar langsung untuk menolak registrasi, menangguhkan layanan akademik, atau menjatuhkan sanksi kepada mahasiswa tanpa melalui verifikasi manual.

Petugas keuangan tetap perlu memvalidasi hasil prediksi dengan merujuk pada riwayat pembayaran, bukti transaksi, program keringanan yang berlaku, komunikasi dengan mahasiswa, dan kebijakan institusi. C4.5 dapat dimanfaatkan sebagai model pembanding atau lapisan penjelasan (*explainability layer*) untuk membantu petugas memahami pola atribut yang mendasari keputusan klasifikasi sistem [6], [12].

Kinerja model juga perlu dievaluasi secara berkala mengingat pola pembayaran dapat berubah seiring perubahan biaya pendidikan, kondisi ekonomi, kebijakan pembayaran bertahap, dan komposisi mahasiswa. Pemantauan direkomendasikan menggunakan metrik *recall*, F1-*score*, *balanced accuracy*, dan PR-AUC dengan mempertahankan definisi kelas yang konsisten [14], [15].

Penggunaan atribut yang mengandung informasi pribadi, seperti alamat, pendidikan wali, pekerjaan wali, dan penghasilan wali, memerlukan pengelolaan data yang berhati-hati dan bertanggung jawab. Akses terhadap data tersebut perlu dibatasi sesuai kewenangan dan tujuan penggunaan. Institusi juga perlu menyediakan mekanisme pemutakhiran data apabila informasi administratif mahasiswa atau wali tidak lagi mencerminkan kondisi aktual.

4.5. Keterbatasan Penelitian

Penelitian ini menggunakan 258 rekaman dari satu perguruan tinggi, sehingga hasilnya belum dapat digeneralisasikan secara langsung ke institusi lain yang memiliki karakteristik mahasiswa, kebijakan penetapan biaya, dan sistem penagihan yang berbeda. Meskipun prosedur RSKF memperkuat validitas internal penelitian [10], prosedur tersebut belum menggantikan validasi eksternal menggunakan data dari institusi lain.

Penelitian ini juga belum menerapkan validasi temporal yang memisahkan data berdasarkan periode waktu. Pengujian temporal diperlukan untuk mengkaji kemampuan model ketika diterapkan pada data piutang dari periode yang lebih baru. Simpangan baku akurasi ANN sebesar 6,17% menunjukkan bahwa kinerja model masih dipengaruhi oleh variasi komposisi data latih dan data uji antar-*fold* [10].

Definisi operasional piutang tertagih dan tak tertagih perlu didokumentasikan secara konsisten untuk memastikan bahwa seluruh prediktor yang digunakan tersedia sebelum status akhir piutang ditetapkan, guna mencegah penggunaan informasi masa depan (*data leakage*) dalam proses prediksi [18], [19], [20], [21].

Analisis interpretabilitas pada penelitian ini tetap memiliki sejumlah keterbatasan metodologis yang perlu dicermati. *Grouped permutation importance* berpotensi menghasilkan nilai kepentingan yang terbagi, tereduksi, atau justru terkonsentrasi pada salah satu atribut tatkala terdapat korelasi yang kuat antarprediktor [22]. Keterbatasan ini relevan dengan konteks penelitian, mengingat Jumlah Piutang UKT Mahasiswa berkorelasi sangat kuat dengan Umur Piutang UKT Mahasiswa, demikian pula hubungan antara Jumlah Piutang BPP dan Umur Piutang BPP. Dengan demikian, peringkat kepentingan atribut yang dihasilkan tidak dapat ditafsirkan sebagai ukuran pengaruh yang sepenuhnya independen, dan juga tidak dapat dimaknai sebagai bukti hubungan kausal antara atribut dan status piutang [12], [13]. Penelitian selanjutnya disarankan untuk melengkapi analisis ini dengan pendekatan *conditional permutation importance*, SHAP (*SHapley Additive exPlanations*), atau *accumulated local effects* (ALE) [22], serta menyertakan validasi temporal dan validasi eksternal guna menguji kestabilan pola kepentingan atribut pada periode waktu dan institusi yang berbeda.

5. KESIMPULAN

Penelitian ini berhasil membandingkan kemampuan algoritma C4.5 [6], *Artificial Neural Network* (ANN) [8], [9], dan Gaussian Naïve Bayes [10] dalam memprediksi piutang biaya pendidikan mahasiswa tak tertagih berdasarkan 258 rekaman mahasiswa. Hasil pengujian *holdout* menunjukkan bahwa kinerja model dipengaruhi oleh komposisi data latih dan data uji yang digunakan. Pada pembagian data 80:20, C4.5 dan ANN memperoleh akurasi yang setara sebesar 76,92%, sedangkan pada pembagian data 90:10 C4.5 menghasilkan akurasi tertinggi sebesar 84,62%. Mengingat data uji pada skenario 90:10 hanya terdiri atas 26 rekaman, hasil evaluasi *RepeatedStratifiedKFold* 5×10 digunakan sebagai dasar utama dalam membandingkan kinerja antarmodel secara lebih representatif [10], [16], [17].

Berdasarkan evaluasi *RepeatedStratifiedKFold* 5×10, ANN memperoleh nilai rata-rata tertinggi pada sebagian besar metrik, yaitu *accuracy* sebesar 82,09% ± 6,17%, *recall* 82,71%, F1-*score* 78,62%, *balanced accuracy* 82,20%, dan PR-AUC 91,32% [14]. C4.5 menghasilkan kinerja yang sangat mendekati ANN, dengan *accuracy* 81,04% ± 4,98%, *precision* 76,71%, *recall* 77,54%, *balanced accuracy* 80,43%, dan PR-AUC 89,08%. C4.5 juga memiliki *precision* yang sedikit lebih tinggi serta simpangan baku akurasi yang lebih kecil dibandingkan ANN, sehingga mencerminkan karakteristik prediksi yang lebih selektif dan relatif lebih stabil antar-*fold* [6]. Adapun Gaussian Naïve Bayes menghasilkan *recall* yang relatif tinggi sebesar 80,58%, namun disertai *accuracy* hanya 48,26% ± 8,90%, *precision* 42,64%, *balanced accuracy* 53,86%, dan PR-AUC 46,99%, yang mengindikasikan kecenderungan model menghasilkan prediksi positif secara berlebihan dan kurang mampu membedakan kelas tertagih dan tak tertagih secara seimbang [14], [15].

ANN ditetapkan sebagai model utama karena tujuan operasional penelitian lebih memprioritaskan kemampuan mendeteksi kasus piutang tak tertagih, sebagaimana tercermin dari nilai *recall*-nya yang lebih tinggi [3], [4]. Meskipun demikian, selisih kinerja antara ANN dan C4.5 tergolong relatif terbatas, yaitu 1,05 poin pada *accuracy*, 1,77 poin pada *balanced accuracy*, dan 2,24 poin pada PR-AUC. Oleh karena pengujian signifikansi statistik terhadap skor kedua model pada *fold* yang sama belum dilakukan, hasil penelitian ini tidak dapat digunakan untuk menyatakan bahwa ANN secara signifikan maupun universal lebih unggul daripada C4.5. C4.5 tetap menjadi alternatif yang kompetitif, khususnya apabila institusi lebih mengutamakan kestabilan kinerja, *precision*, dan keterpahaman aturan klasifikasi [6], [7].

Analisis *grouped permutation importance* [12], [13], [25], [26] menunjukkan bahwa Jumlah Piutang UKT Mahasiswa merupakan atribut prediktif paling berpengaruh secara konsisten pada ketiga algoritma. Pada C4.5, pengacakan atribut tersebut menurunkan PR-AUC rata-rata sebesar 46,85 poin dan *balanced accuracy* sebesar 28,82 poin pada seluruh *fold*, yang mengindikasikan ketergantungan model yang sangat terkonsentrasi pada atribut ini [6], [12]. Pada ANN, kontribusi prediktif tersebar lebih merata antara Jumlah Piutang UKT Mahasiswa, Umur Piutang UKT Mahasiswa, Jumlah Piutang BPP, dan Umur Piutang BPP [8], [9]. Gaussian Naïve Bayes juga menempatkan jumlah dan umur piutang UKT sebagai dua atribut utama, meskipun dengan pengaruh yang lebih kecil dan kurang

https://doi.org/10.25077/TEKNOSI.v12i2.2026.312-326

stabil dibandingkan kedua model lainnya [22]. Sementara itu, atribut akademik, wilayah, pendidikan wali, pekerjaan wali, dan penghasilan wali memberikan kontribusi prediktif tambahan yang relatif kecil atau tidak konsisten setelah atribut jumlah dan umur piutang diperhitungkan dalam model [12], [13].

Perbandingan distribusi kelas memperlihatkan bahwa rekaman piutang tak tertagih memiliki median jumlah dan umur piutang yang lebih tinggi dibandingkan rekaman piutang tertagih. Perbedaan paling besar terdapat pada Jumlah Piutang UKT Mahasiswa, dengan median sebesar Rp3.000.000 pada kelas tak tertagih berbanding Rp0 pada kelas tertagih. Temuan ini mengindikasikan bahwa peningkatan jumlah dan umur piutang berasosiasi dengan meningkatnya kemungkinan suatu rekaman diklasifikasikan sebagai piutang tak tertagih. Namun, hasil interpretasi tersebut merupakan hubungan prediktif (asosiatif) dan tidak dapat dimaknai sebagai hubungan sebab-akibat (kausal) [12], [13]. Validitas interpretasi ini juga mensyaratkan bahwa atribut jumlah dan umur piutang telah tersedia sebelum status akhir piutang ditetapkan, guna mencegah terjadinya kebocoran target (*target leakage*) [18], [21].

Kontribusi penelitian ini terhadap pengembangan ilmu pengetahuan dan teknologi tidak terletak pada penciptaan algoritma klasifikasi baru, melainkan pada pengembangan kerangka evaluasi dan interpretabilitas model prediksi piutang pendidikan yang lebih seragam, transparan, dan berorientasi pada kelas prioritas. Kerangka tersebut mengintegrasikan penggunaan *fold* yang identik antarmodel, penempatan tahap imputasi, *encoding*, standardisasi, dan *oversampling* di dalam *pipeline* pelatihan [18], [19], [20], evaluasi *RepeatedStratifiedKFold* [10], [16], [17], penggunaan metrik *balanced accuracy* dan PR-AUC [14], [15], serta penerapan *grouped permutation importance* secara konsisten pada tiga algoritma yang memiliki struktur internal berbeda [12], [13]. Pendekatan ini memungkinkan perbandingan kinerja dan kepentingan atribut dilakukan secara lebih adil dibandingkan penggunaan ukuran kepentingan internal yang berbeda-beda pada setiap model.

Secara aplikatif, hasil penelitian ini memberikan dasar ilmiah bagi pengembangan sistem pendukung keputusan untuk membantu petugas keuangan menyusun prioritas pemeriksaan piutang berdasarkan tingkat risiko prediktif [3], [4], [23], [24] . Jumlah dan umur piutang UKT dapat dijadikan indikator utama, sedangkan informasi piutang BPP dapat berfungsi sebagai indikator pelengkap. Hasil prediksi tetap harus diperlakukan sebagai rekomendasi analitik yang memerlukan verifikasi manual, dan tidak boleh digunakan secara langsung untuk menolak registrasi, membatasi layanan akademik, atau menjatuhkan sanksi kepada mahasiswa. Dengan demikian, seluruh tujuan penelitian telah tercapai, yakni menghasilkan prediksi menggunakan ketiga algoritma, mengidentifikasi atribut yang paling berpengaruh secara konsisten, serta memberikan pemahaman komparatif mengenai kelebihan, keterbatasan, dan implikasi operasional dari masing-masing model.

DAFTAR PUSTAKA

[1] P. A. Prayesy and E. S. Negara, "Classification of the Fluency Multipurpose of Bank Mandiri Credit Payments Based on Debtor Preferences Using Naive Bayes and Neural Network Method," vol. 7, no. 1, pp. 7–16, Jun. 2022, doi: 10.15575/join.v7i1.762.

[2] I. T. A. Nur, N. Y. Setiawan, and F. A. Bachtiar, "Perbandingan Performa Metode Klasifikasi SVM, Neural Network, dan Naive Bayes untuk Mendeteksi Kualitas Pengajuan Kredit di Koperasi Simpan Pinjam," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 6, no. 4, p. 444, Jul. 2019, doi: 10.25126/jtiik.2019641352.

[3] X. Zhu, Q. Chu, X. Song, P. Hu, and L. Peng, "Explainable prediction of loan default based on machine learning models," *Data Sci. Manag.*, vol. 6, no. 3, pp. 123–133, Sep. 2023, doi: 10.1016/j.dsm.2023.04.003.

[4] X. Zhang *et al.*, "Data-Driven Loan Default Prediction: A Machine Learning Approach for Enhancing Business Process Management," *Systems*, vol. 13, no. 7, p. 581, Jul. 2025, doi: 10.3390/systems13070581.

[5] V. S. Ginting, K. Kusrini, and E. T. Luthfi, "Penerapan Algoritma C4.5 Dalam Memprediksi Keterlambatan Pembayaran Uang Sekolah Menggunakan Python," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 1–6, Jun. 2020, doi: 10.36294/jurti.v4i1.1101.

[6] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann Publishers, 1993.

[7] J. R. Quinlan, "Induction of Decision Trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986, doi: 10.1007/BF00116251.

[8] G. E. Hinton, "Connectionist Learning Procedures," *Artif. Intell.*, vol. 40, no. 1–3, pp. 185–234, 1989, doi: 10.1016/0004-3702(89)90049-0.

[9] scikit-learn developers, "1.17. Neural Network Models (Supervised) — MLPClassifier." 2024. Accessed: Jun. 21, 2026. [Online]. Available: https://scikit-learn.org/stable/modules/neural_networks_supervised.html

[10] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011, doi: 10.5555/1953048.2078195.

[11] A. Rahmayanti, L. Rusdiana, and S. Suratno, "Perbandingan Metode Algoritma C4.5 Dan Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa," *Walisongo J. Inf. Technol.*, vol. 4, no. 1, pp. 11–22, Aug. 2022, doi: 10.21580/wjit.2022.4.1.9654.

[12] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[13] B. Gregorutti, B. Michel, and P. Saint-Pierre, "Grouped Variable Importance with Random Forests and Application to Multiple Functional Data Analysis," *Comput. Stat. Data Anal.*, vol. 90, pp. 15–35, 2015, doi: 10.1016/j.csda.2015.04.002.

[14] J. Davis and M. Goadrich, "The Relationship Between Precision-Recall and ROC Curves," in *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, New York, NY, USA: ACM, 2006, pp. 233–240. doi: 10.1145/1143844.1143874.

[15] M. Imani, M. Joudaki, A. Bagheri, and H. R. Arabnia, "Why ROC-AUC Is Misleading for Highly Imbalanced Data: In-Depth Evaluation of MCC, F2-Score, H-Measure, and AUC-Based Metrics Across Diverse Classifiers," *Technologies*, vol. 14, no. 1, p. 54, Jan. 2026, doi: 10.3390/technologies14010054.

[16] S. Szeghalmy and A. Fazekas, "A Comparative Study of the Use of Stratified Cross-Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced

Learning," *Sensors*, vol. 23, no. 4, p. 2333, Feb. 2023, doi: 10.3390/s23042333.

[17] V. Lumumba, D. Kiprotich, M. Mpaine, N. Makena, and M. Kavita, "Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models," *Am. J. Theor. Appl. Stat.*, vol. 13, no. 5, pp. 127–137, Oct. 2024, doi: 10.11648/j.ajtas.20241305.13.

[18] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," *J. Mach. Learn. Res.*, vol. 18, no. 17, pp. 1–5, 2017.

[19] M. Rosenblatt, L. Tejavibulya, R. Jiang, S. Noble, and D. Scheinost, "Data leakage inflates prediction performance in connectome-based machine learning models," *Nat. Commun.*, vol. 15, no. 1, p. 1829, Feb. 2024, doi: 10.1038/s41467-024-46150-w.

[20] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, p. 100804, Sep. 2023, doi: 10.1016/j.patter.2023.100804.

[21] F. R. Pratikto, "Oversampling Sintetis Berbasis Kopula untuk Model Klasifikasi dengan Data yang Tidak Seimbang," *J. Rekayasa Sist. Ind.*, vol. 12, no. 1, pp. 1–10, Apr. 2023, doi: 10.26593/jrsi.v12i1.6380.1-10.

[22] K. K. Nicodemus, J. D. Malley, C. Strobl, and A. Ziegler, "The Behaviour of Random Forest Permutation-Based Variable Importance Measures Under Predictor Correlation," *BMC Bioinformatics*, vol. 11, p. 110, 2010, doi: 10.1186/1471-2105-11-110.

[23] S. Mestiri, "Credit scoring using machine learning and deep Learning-Based models," *Data Sci. Finance Econ.*, vol. 4, no. 2, pp. 236–248, 2024, doi: 10.3934/DSFE.2024009.

[24] T. T. Do, G. Babaei, and P. Pagnottoni, "Explainable Machine Learning for Credit Risk Management When Features are Dependent," *Meas. Interdiscip. Res. Perspect.*, vol. 22, no. 4, pp. 315–340, Oct. 2024, doi: 10.1080/15366367.2023.2261186.

[25] C. Molnar, G. König, B. Bischl, and G. Casalicchio, "Model-agnostic feature importance and effects with dependent features: a conditional subgroup approach," *Data Min. Knowl. Discov.*, vol. 38, no. 5, pp. 2903–2941, Sep. 2024, doi: 10.1007/s10618-022-00901-9.

[26] I. M. Zubair, Y.-S. Lee, and B. Kim, "A New Permutation-Based Method for Ranking and Selecting Group Features in Multiclass Classification," *Appl. Sci.*, vol. 14, no. 8, p. 3156, Apr. 2024, doi: 10.3390/app14083156.