

Model Prediktif Keterlambatan Pembayaran Mahasiswa Berbasis Seleksi Fitur dengan Particle Swarm Optimization

Yessica Fara Desvia^{1*}, Suharjanti², Suhardjono³, Irmawati Carolina³, Resti Lia Andharsaputri³

¹ Politeknik Jatiluhur, Bisnis Digital, Purwakarta, Indonesia

² Teknologi Informasi, Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia

³ Sistem Informasi, Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Email: ^{1*}yessicadesvia@polijati.ac.id, ²suharjanti.shj@bsi.ac.id, ³suhardjono@bsi.ac.id, ⁴irmawati.imc@bsi.ac.id,

⁵resti.ria@bsi.ac.id

Email Penulis Korespondensi: yessicadesvia@polijati.ac.id*

Submitted: 01/08/2025; Accepted: 06/09/2025; Published: 31/12/2025

Abstrak– Keterlambatan pembayaran biaya kuliah menjadi salah satu permasalahan krusial di perguruan tinggi swasta yang dapat berdampak pada risiko akademik, seperti cuti atau putus studi. Penelitian ini diarahkan untuk mengembangkan model prediktif dalam mengidentifikasi keterlambatan pembayaran oleh mahasiswa, dengan memanfaatkan algoritma klasifikasi *Decision Tree* dan *Random Tree*, serta menerapkan metode *Particle Swarm Optimization (PSO)* untuk proses seleksi fitur. Data yang digunakan dalam penelitian ini mencakup 15.697 mahasiswa, masing-masing memiliki enam atribut sebagai variabel prediktor serta satu atribut target yang menunjukkan status mahasiswa, yaitu aktif atau cuti. Tahapan penelitian mencakup pengumpulan data, pra-pemrosesan, klasifikasi, seleksi fitur, dan evaluasi model dilakukan dengan menggunakan metrik akurasi, serta kurva *ROC* dan nilai *AUC*. Hasil penelitian menunjukkan akurasi model mencapai 98,83%, dengan peningkatan signifikan *AUC* pada *Random Tree* dari 0,632 menjadi 0,825 setelah seleksi fitur menggunakan *PSO*. Temuan ini menunjukkan bahwa *PSO* efektif dalam meningkatkan performa model klasifikasi dan mengurangi kompleksitas fitur yang tidak relevan. Sistem prediktif yang dihasilkan dapat membantu institusi pendidikan dalam melakukan deteksi dini mahasiswa berisiko menunggak, sehingga memungkinkan pengambilan tindakan preventif dan intervensi lebih tepat sasaran untuk mendukung keberlangsungan akademik mahasiswa.

Kata Kunci: *Decision Tree*; Keterlambatan Pembayaran; Klasifikasi; *PSO*; *Random Tree*.

Abstract–Late payment of tuition fees is a crucial issue in private universities, potentially leading to academic risks such as student leave or dropout. This study aims to develop a predictive model for identifying late payments by students, utilizing *Decision Tree* and *Random Tree* classification algorithms, and applying the *Particle Swarm Optimization (PSO)* method for feature selection. The data used in this study includes 15,697 students, each with six attributes as predictor variables and one target attribute indicating student status, whether active or on leave. The research stages include data collection, pre-processing, classification, feature selection, and model evaluation using accuracy metrics, ROC curves, and *AUC* values. The results showed that the model accuracy reached 98.83%, with a significant increase in *AUC* on *Random Tree* from 0.632 to 0.825 after feature selection using *PSO*. These findings indicate that *PSO* is effective in improving the performance of the classification model and reducing the complexity of irrelevant features. The resulting predictive system can help educational institutions in early detection of students at risk of delinquency, thus enabling more targeted preventive and intervention actions to support student academic sustainability.

Keywords: Classification; *Decision Tree*; Late Payment; *PSO*; *Random Tree*.

1. PENDAHULUAN

Pendidikan tinggi swasta di Indonesia menghadapi tantangan dalam menjaga keberlangsungan studi mahasiswa, terutama terkait keterlambatan pembayaran biaya kuliah [1], [2]. Keterlambatan ini sering dipicu oleh faktor ekonomi keluarga, kurangnya pemahaman terhadap jadwal pembayaran, serta minimnya sistem peringatan dini dari pihak kampus [3], [4]. Kondisi tersebut tidak hanya berisiko menyebabkan mahasiswa mengambil cuti atau putus studi, tetapi juga berdampak pada stabilitas keuangan dan reputasi institusi [5], [6]. Banyak perguruan tinggi masih bersifat reaktif karena belum memiliki sistem prediktif untuk mengidentifikasi mahasiswa yang berpotensi menunggak sebelum jatuh tempo [7], [8].

Pendekatan berbasis data dan algoritma klasifikasi memiliki potensi untuk melakukan deteksi dini terhadap mahasiswa yang berisiko menunggak, termasuk mempertimbangkan faktor sosial-ekonomi dan variabel relevan lainnya [9], [10]. Evaluasi fitur dan pemilihan atribut yang tepat menjadi penting agar model prediksi tidak hanya akurat tetapi juga efisien dan interpretatif [11], [12].

Sejumlah penelitian sebelumnya telah mencoba menghadirkan solusi prediktif dengan berbagai algoritma klasifikasi. Pendekatan berbasis *Random Forest* terbukti efektif dalam memprediksi kelancaran pembayaran mahasiswa berdasarkan data historis [13]. Studi lain yang membandingkan *Naive Bayes* dan *K-Nearest Neighbor (KNN)* menemukan bahwa *KNN* memiliki akurasi lebih tinggi, meskipun seluruh fitur digunakan tanpa seleksi [14]. *Decision Tree C4.5* pernah digunakan untuk memprediksi mahasiswa yang berisiko cuti studi, tetapi tidak diarahkan secara langsung pada keterlambatan pembayaran [15]. Kombinasi *Naive Bayes* dan *SVM* juga diterapkan, dengan *SVM* lebih unggul namun tanpa optimasi atau seleksi fitur [16].

Beberapa studi lebih komprehensif membandingkan lima algoritma populer (*KNN*, *Naive Bayes*, *C4.5*, *Random Forest*, *Logistic Regression*) dan menggunakan metode seleksi fitur seperti *Mutual Information*, *Forward*

Selection, *Backward Elimination*, dan *Recursive Feature Elimination*. Meski demikian, studi tersebut tidak memanfaatkan *Particle Swarm Optimization (PSO)* dan tidak secara spesifik menyoroti keterlambatan pembayaran mahasiswa [17]. Ada penelitian lain yang hanya menggunakan *KNN*, tanpa seleksi fitur atau perbandingan dengan algoritma lain [18]. Pendekatan *SVM+PSO* menunjukkan akurasi meningkat, tetapi model kompleks dan kurang mudah diinterpretasikan bagi pihak administrasi kampus [19].

Selain literatur teknis, perspektif manajemen pendidikan dan keuangan menunjukkan pentingnya pengelolaan risiko keterlambatan pembayaran di perguruan tinggi swasta [20]. Dalam konteks ini, sistem prediksi keterlambatan pembayaran mahasiswa dapat berfungsi sebagai alat strategis untuk mengetahui mahasiswa yang terlambat dalam pembayaran. Analisis pembiayaan dan manajemen finansial di perguruan tinggi swasta sangat penting untuk meningkatkan keberlanjutan finansial dan kualitas pendidikan [21]. Dengan demikian, integrasi pendekatan data mining dengan manajemen keuangan perguruan tinggi memperkuat relevansi penelitian ini.

Dari kajian literatur tersebut, masih terdapat beberapa celah yang perlu diatasi. Belum ada penelitian yang secara khusus mengintegrasikan algoritma *Decision Tree (DT)* dan *Random Tree* dengan seleksi fitur berbasis *PSO* untuk prediksi keterlambatan pembayaran mahasiswa. Sebagian besar studi sebelumnya menitikberatkan pada akurasi tanpa mempertimbangkan efisiensi dan interpretabilitas, sehingga model sulit diaplikasikan secara praktis. Selain itu, hubungan antara hasil prediksi dan strategi preventif, seperti pemberian peringatan dini atau kebijakan keuangan adaptif, masih jarang dieksplorasi secara sistematis.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi keterlambatan pembayaran mahasiswa yang akurat, efisien, dan mudah diinterpretasikan dengan menggunakan *DT* dan *Random Tree*, serta dukungan seleksi fitur berbasis *PSO*. Model ini diharapkan mampu mengidentifikasi mahasiswa yang berisiko menunggak pembayaran dan memberikan sinyal peringatan dini bagi pihak kampus dalam merancang strategi preventif, seperti pemberian subsidi internal, modifikasi skema cicilan, atau konsultasi finansial.

Kontribusi penelitian terletak pada pengembangan model yang mengintegrasikan algoritma pohon keputusan baik *DT* maupun *Random Tree* dengan *PSO* secara inovatif, menawarkan pendekatan yang lebih interpretatif dibandingkan metode kompleks seperti *SVM*, dan memberikan manfaat praktis berupa sistem prediksi yang mendukung pengambilan keputusan berbasis data dalam manajemen akademik. Penelitian ini menghadirkan solusi yang belum pernah diimplementasikan sebelumnya di konteks perguruan tinggi swasta di Indonesia.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metode penelitian ini mengikuti tahapan yang sistematis untuk membangun model prediksi keterlambatan pembayaran mahasiswa guna mencegah risiko cuti atau keluar studi. Alur penelitian terdiri dari beberapa langkah utama: pengumpulan data, pra-pemrosesan data, proses klasifikasi dilakukan dengan memanfaatkan algoritma pohon keputusan seperti (*Decision Tree* dan *Random Tree*), seleksi fitur menggunakan *PSO*, dan evaluasi model. Gambar 1 menggambarkan alur penelitian secara umum.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini dikumpulkan dari sistem informasi akademik milik salah satu perguruan tinggi swasta di Indonesia. Dataset terdiri dari 15.697 data mahasiswa, yang mencakup status administrasi keuangan dan status akademik. Dataset ini bersifat tabular dan mencakup tujuh atribut, yaitu enam atribut reguler dan satu atribut spesial.

Atribut reguler yang digunakan adalah: 1). Status Aktif/Cuti bahwa mahasiswa pernah melakukan cuti atau tidak, 2). Semester: posisi semester aktif saat data diambil, 3). Jenis kelamin: laki-laki atau perempuan, 4) Jurusan SLTA: latar belakang jurusan sekolah menengah atas, 5). Kuliah sambil bekerja: status bekerja sambil kuliah (ya/tidak), 6). Pekerjaan ayah: informasi pekerjaan orang tua sebagai variabel sosial ekonomi.

Sedangkan atribut spesial adalah: Nomor biaya kuliah, target klasifikasi biner yang menunjukkan apakah mahasiswa menunggak pembayaran kuliah atau tidak menunggak.

Seluruh data telah dianonimkan untuk menjaga kerahasiaan identitas mahasiswa dan memenuhi etika penelitian.

2.3 Pra-Pemrosesan Data

Sebelum digunakan untuk pelatihan model, data perlu melalui beberapa tahap pra-pemrosesan, meliputi: 1). Pemeriksaan nilai kosong: memastikan tidak ada missing values yang dapat mempengaruhi hasil klasifikasi, 2). Transformasi data kategorikal: Seluruh atribut seperti jenis kelamin, jurusan SLTA, kuliah sambil bekerja, status pekerjaan ayah diubah ke dalam bentuk numerik menggunakan teknik one-hot encoding, 3). Penghapusan duplikasi: data, dan status aktif cuti yang memiliki entri duplikat dihapus agar tidak mempengaruhi proses pembelajaran model.

Tahapan ini bertujuan untuk meningkatkan kualitas data dan meminimalkan noise yang dapat mengurangi akurasi prediksi.

2.3 Klasifikasi

Model klasifikasi dibangun memanfaatkan dua algoritma pohon keputusan, yaitu *Decision Tree* dan *Random Tree*. Keduanya digunakan untuk membandingkan kinerja model berbasis rule dengan interpretabilitas tinggi.

Algoritma *Decision Tree* merupakan metode klasifikasi yang memetakan data ke dalam struktur pohon bercabang berdasarkan atribut-atribut yang paling informatif. Tujuannya adalah memaksimalkan "pemisahan" antar kelas. *Decision Tree* digunakan sebagai baseline, dengan parameter default dari pustaka RapidMiner. Langkah-langkah *Decision Tree*: 1). Hitung entropy dataset awal, 2). Hitung gain untuk setiap atribut, 3). Pilih atribut dengan gain tertinggi sebagai node, dengan Information Gain tertinggi akan menjadi simpul (node) pada pohon, 4). Pisahkan dataset berdasarkan atribut tersebut, dengan setiap cabang merepresentasikan satu nilai dari atribut yang dipilih, 5). Ulangi langkah 1–4 pada setiap subset hingga seluruh data dalam subset tersebut memiliki label seragam atau tidak terdapat atribut yang tersisa.

Random Tree varian dari *Decision Tree* yang melakukan pemilihan atribut secara acak dari subset atribut yang tersedia. Pendekatan ini meningkatkan keragaman antar pohon ketika digunakan dalam ensemble, sehingga membangun banyak pohon secara acak dan menggabungkannya untuk menghasilkan prediksi akhir, lebih tahan terhadap overfitting. Langkah-langkah Random Tree: 1). Ambil sampel acak dari dataset (tanpa penggantian), pengambilan bisa seluruh data atau sebagian, 2). Pada tiap node, pilih subset acak dari atribut, 3). Hitung gain atau kriteria split lain pada atribut-atribut acak tersebut, 4). Pilih atribut terbaik dari subset acak untuk split node, 5). Bangun node dan ulangi rekursif hingga kondisi berhenti terpenuhi, 6). Prediksi dilakukan dengan traversing dari root ke daun berdasarkan atribut input.

Model dilatih menggunakan data 80% untuk training dan 20% untuk testing. Seluruh eksperimen dilakukan menggunakan perangkat lunak RapidMiner Studio.

2.4 Seleksi Fitur dengan PSO

Untuk mengurangi kompleksitas model dan meningkatkan akurasi klasifikasi, dilakukan seleksi fitur menggunakan algoritma optimasi *PSO*. *PSO* digunakan untuk mencari subset fitur paling relevan yang memiliki kontribusi signifikan terhadap prediksi status aktif atau cuti [22].

Setiap partikel dalam *PSO* mewakili kombinasi fitur biner (0 untuk tidak dipilih, 1 untuk dipilih). Fungsi objektif diukur berdasarkan akurasi validasi silang dari model klasifikasi yang dibangun menggunakan fitur tersebut. Parameter *PSO* diatur sebagai berikut: Ukuran populasi: 5 partikel, dan Iterasi maksimum: 30 untuk menjaga efisiensi dan mempertimbangkan keterbatasan komputasi, sekaligus tetap menghasilkan model yang stabil. Proses ini diharapkan dapat menghasilkan kombinasi fitur terbaik yang selanjutnya digunakan untuk pelatihan model.

2.5 Evaluasi Model

Proses evaluasi model memanfaatkan empat metrik utama
Akurasi: proporsi prediksi benar terhadap seluruh data dengan rumus (1)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision: kemampuan model mengklasifikasikan mahasiswa benar-benar cuti dari seluruh prediksi cuti dengan rumus (2)

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall: kemampuan model mendeteksi seluruh mahasiswa yang benar-benar cuti dengan rumus (3)

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Keterangan:

TP = True Positive

TN = True Negative

FP = *False Positive*

FN = *False Negative*

Selain metrik lainnya, analisis juga melibatkan kurva *ROC (Receiver Operating Characteristic)* serta penghitungan nilai *AUC (Area Under Curve)*. untuk menilai performa klasifikasi secara menyeluruh. Evaluasi dilakukan pada model sebelum dan sesudah seleksi fitur *PSO* untuk menilai peningkatan performa prediksi akibat pengurangan fitur. Seluruh eksperimen dijalankan menggunakan RapidMiner Studio 5.0 pada komputer dengan prosesor Intel i5, RAM 8 GB, dan sistem operasi Windows 10

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang diperoleh terdiri dari satu atribut sebagai target klasifikasi, yakni biaya kuliah menunggak atau tidak menunggak, dan enam atribut lain sebagai variabel prediktor. Contoh sebagian data yang telah dipersiapkan untuk analisis dapat dilihat pada Gambar 2.

ExampleSet (15697 examples, 1 special attribute, 6 regular attributes)							
Row No.	no_by_klh	smt	jen_kel	jrs_slta	klh_smbil_krj	pkj_ayh	sts_aktif_cuti
1	1	8	1	1	1	18	0
2	1	8	1	1	1	18	0
3	1	6	0	0	1	18	0
4	1	6	1	0	1	18	0
5	1	6	0	0	1	18	0
6	1	6	0	0	1	18	0
7	1	6	1	0	1	18	0
8	0	4	1	0	1	18	1
9	0	4	0	0	1	18	1
10	1	6	0	0	1	18	0
11	1	6	1	0	1	18	0
12	1	6	0	0	1	18	0
13	1	6	0	0	1	18	0
14	1	6	1	0	1	18	0
15	1	6	1	0	1	18	0
16	1	4	1	0	1	18	0
17	1	2	1	0	1	16	0
18	1	2	0	0	1	16	0
19	1	2	0	0	0	16	0
20	1	2	0	0	0	16	0
21	1	2	0	0	1	18	0
22	1	2	0	0	0	16	0
23	1	2	0	0	0	16	0

Gambar 2. Dataset Pembayaran

Secara umum Gambar 2, data yang diperoleh mencerminkan kondisi aktual mahasiswa dari berbagai latar belakang. Data yang terekam mencakup informasi seperti semester berjalan, jenis kelamin, jurusan asal saat di sekolah menengah, status kuliah sambil bekerja, pernah cuti atau tidak, serta pekerjaan ayah. Tiap entri dalam dataset menggambarkan satu individu mahasiswa dengan karakteristik unik berdasarkan atribut-atribut tersebut.

Dari hasil eksplorasi awal terhadap tabel, terlihat bahwa nilai-nilai dari setiap atribut menunjukkan keragaman isian. Atribut semester mencakup berbagai tingkat semester, mulai dari mahasiswa awal hingga mahasiswa tingkat lanjut. Atribut jenis kelamin dan jurusan SLTA juga memuat nilai-nilai yang merepresentasikan perbedaan latar belakang mahasiswa. Begitu pula dengan atribut pekerjaan ayah, yang menunjukkan variasi pekerjaan yang cukup luas. Status aktif cuti yang menunjukkan apakah mahasiswa dalam kondisi aktif mengikuti perkuliahan atau sedang mengambil cuti akademik

Atribut target pada dataset ini adalah biaya kuliah dimana 0 menerangkan bahwa mahasiswa menunggak bayar, dan 1 mahasiswa tidak menunggak. Status ini menjadi fokus utama dalam proses klasifikasi yang dilakukan pada tahap-tahap selanjutnya.

Data ditampilkan dalam format numerik yang telah disesuaikan sebelumnya agar dapat digunakan dalam proses pemodelan. Format ini memudahkan proses pra-pemrosesan dan pengolahan lebih lanjut di tahap klasifikasi serta seleksi fitur.

3.2 Pra-Pemrosesan Data

Sebelum digunakan dalam proses klasifikasi, data yang terdapat pada Gambar 2 telah melalui tahapan pra-pemrosesan data untuk memastikan kesiapan dan kesesuaiannya dalam proses pemodelan. Tahapan ini penting dilakukan guna menghindari gangguan pada hasil analisis akibat kualitas data yang rendah atau tidak seragam. Langkah awal dalam pra-pemrosesan adalah pengecekan data kosong (*missing values*) dan penghapusan duplikasi. Data yang tidak lengkap atau terduplikasi diidentifikasi dan dihapus agar tidak mempengaruhi hasil model secara

negatif. Selanjutnya, dilakukan transformasi terhadap atribut kategorikal menjadi bentuk numerik. Beberapa atribut yang sebelumnya berupa teks seperti jenis kelamin (*jen_kel*), jurusan SLTA (*jrs_slta*), kuliah sambil bekerja, dan pekerjaan ayah, dikonversi menggunakan teknik pengkodean (*encoding*), agar dapat diproses lebih lanjut oleh algoritma klasifikasi. Transformasi ini diperlukan karena sebagian besar algoritma klasifikasi dalam perangkat lunak RapidMiner tidak dapat memproses data dalam bentuk string atau teks langsung.

Sebagai ilustrasi, Tabel 1 adalah bagian dari data awal yang ditampilkan sebelum dilakukan transformasi.

Tabel 1. Dataset Sebelum Pra-Pemrosesan

Jenis kelamin	Jurusan SLTA	Bekerja	Pekerjaan Ayah
1	IPA	Tidak	Petani
Laki-Laki	TEKNIK KOMPUTER DAN JARINGAN	Ya	Wiraswasta
Perempuan	REKAYASA PERANGKAT LUNAK	Tidak	Karyawan Swasta
Laki-Laki	TKJ	Tidak	Tidak bekerja
Laki-Laki	REKAYASA PERANGKAT LUNAK	Tidak	Pedagang Kecil
Perempuan	IPA	Tidak	Wiraswasta

Seluruh nilai kategorikal tersebut kemudian dikodekan ke dalam angka untuk mempermudah proses komputasi. Misalnya, kategori "Laki-Laki" bisa dikodekan menjadi 1 dan "Perempuan" menjadi 0, atau pekerjaan dikonversi ke dalam bentuk numerik sesuai kategorinya. Selain itu, semua atribut yang digunakan telah distandarkan ke dalam format numerik yang seragam dan konsisten, termasuk normalisasi skala jika diperlukan. Proses ini memastikan bahwa perbedaan skala antar fitur tidak memengaruhi hasil model klasifikasi. Dengan hasil transformasi pada tabel 2.

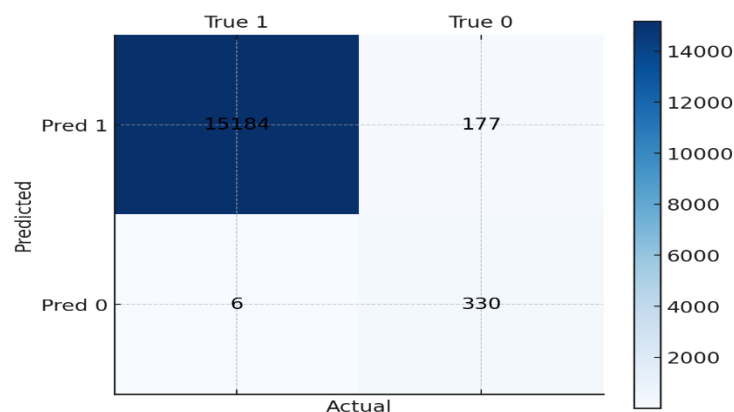
Tabel 2. Dataset Setelah Pra-Pemrosesan

Jenis kelamin	Jurusan SLTA	Bekerja	Pekerjaan Ayah
1	1	0	9
1	1	1	1
0	1	0	3
1	1	0	0
1	1	0	5
0	1	1	2

Melalui tahapan pra-pemrosesan ini, kualitas data ditingkatkan dan siap untuk digunakan dalam proses klasifikasi serta seleksi fitur dengan penjelasan lebih lanjut akan disampaikan pada bagian berikutnya

3.3 Klasifikasi

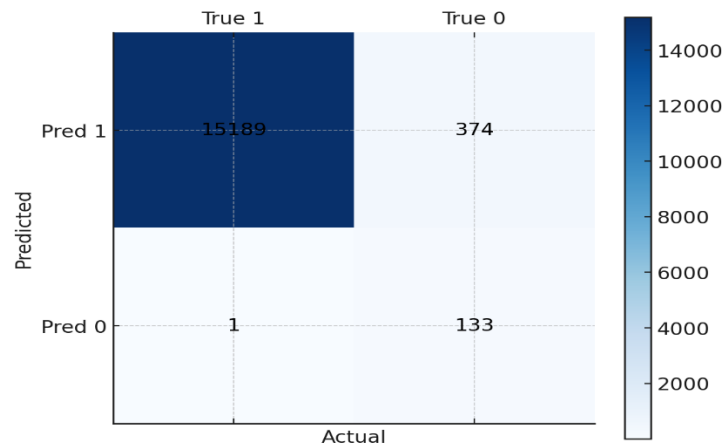
Hasil yang didapat setelah proses klasifikasi menggunakan algoritma *Decision Tree* ditunjukkan pada Gambar 3 dalam bentuk confusion matrix. Matriks ini memperlihatkan performa model dalam mengklasifikasikan status keterlambatan pembayaran mahasiswa berdasarkan data yang telah dipra-proses.



Gambar 3. Confusion Matrix *Decision Tree*

Pada Gambar 3 dihasilkan TP sebanyak 15.184 mahasiswa diprediksi tidak menunggak dan memang benar tidak menunggak. FP sebanyak 177 mahasiswa diprediksi tidak menunggak, padahal menunggak. FN sebanyak 6 mahasiswa diprediksi menunggak, padahal tidak menunggak. TN sebanyak 330 mahasiswa diprediksi tidak menunggak dan memang tidak menunggak. Hasil dari confusion matrix tersebut dapat dihitung akurasi dengan rumus (1), *precision* dengan rumus (2), dan *recall* dengan rumus (3) sehingga menghasilkan akurasi sebesar 98,83%, *precision* 98,21%, *recall* 65,09%.

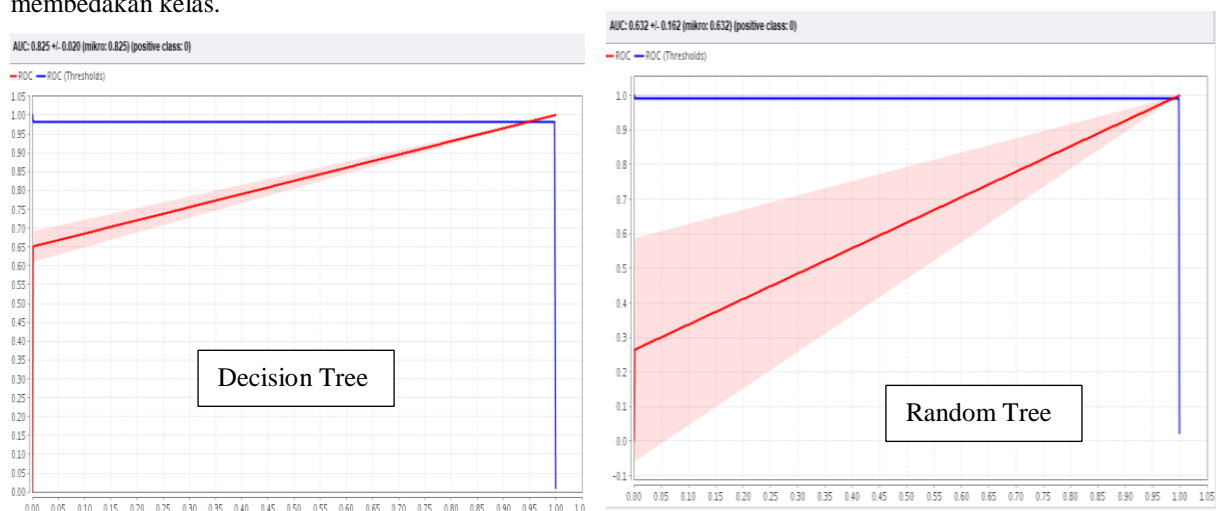
Sementara itu, hasil klasifikasi menggunakan algoritma Random Tree dapat dilihat pada Gambar 4. Confusion matrix ini menunjukkan akurasi dan kemampuan model dalam membedakan mahasiswa yang berisiko menunggak dan tidak.



Gambar 4. Confusion Matrix Random Tree

Pada Gambar 4 dihasilkan TP sebanyak 15.189 mahasiswa diprediksi tidak menunggak dan benar tidak menunggak. FP sebanyak 374 mahasiswa diprediksi tidak menunggak, tapi ternyata menunggak. FN sebanyak 1 mahasiswa diprediksi menunggak, padahal tidak menunggak. TN sebanyak 133 mahasiswa diprediksi menunggak dan memang menunggak. Hasil dari confusion matrix tersebut dapat dihitung akurasi dengan rumus (1), *precision* dengan rumus (2), dan *recall* dengan rumus (3) sehingga menghasilkan akurasi sebesar 97,61%, *precision* 99,25%, *recall* 26,33%.

Setelah menghasilkan confusion matrix, kedua model juga menghasilkan gambar kurva ROC untuk model prediktif keterlambatan pembayaran mahasiswa yang dibangun menggunakan algoritma *Decision Tree* dan Random Tree. Kurva ROC dimanfaatkan sebagai alat evaluasi performa model klasifikasi dengan menganalisis hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada beragam nilai ambang batas (threshold). Semakin dekat posisi kurva ROC ke sudut kiri atas, maka semakin tinggi kinerja model dalam membedakan kelas dengan tepat. Nilai *AUC* digunakan sebagai indikator kuantitatif akurasi model dalam membedakan kelas.



Gambar 5. Kurva ROC Decision & Random Tree

Gambar 5 menjelaskan *AUC* sebesar 0.825 memberitahukan model *Decision Tree* cukup baik dalam membedakan antara mahasiswa yang membayar tepat waktu dan yang terlambat. *ROC curve* berada jauh di atas garis diagonal menunjukkan kemampuan klasifikasi yang kuat. Variasi error kecil (± 0.020) menandakan model ini lebih stabil dan andal. *AUC* sebesar 0.632 menunjukkan bahwa model Random Tree kurang mampu

membedakan antara mahasiswa yang tepat waktu dan terlambat dalam membayar. ROC-nya mendekati garis diagonal, artinya model hampir acak dalam klasifikasi.

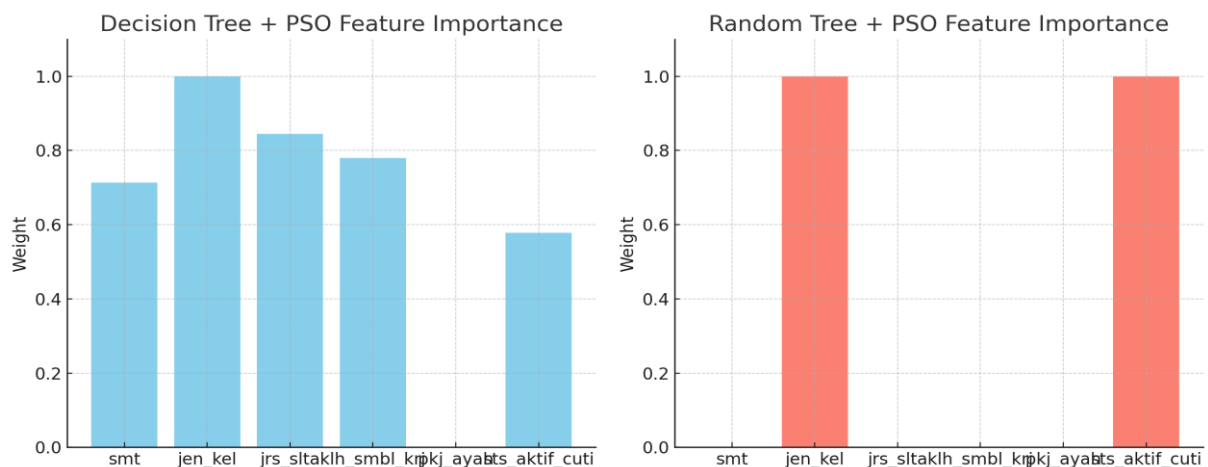
3.4 Seleksi Fitur dengan PSO

Untuk meningkatkan akurasi model klasifikasi, dilakukan proses seleksi dan pembobotan atribut menggunakan metode *PSO*. *PSO* berfungsi untuk mengidentifikasi atribut-atribut yang paling relevan dalam memprediksi keterlambatan pembayaran mahasiswa dengan cara mengoptimalkan bobot masing-masing fitur. Atribut yang memiliki bobot dengan nilai tinggi dianggap memiliki kontribusi yang lebih besar terhadap prediksi akhir, sementara atribut yang memiliki bobot rendah atau nol dapat dianggap tidak relevan. Hasil pembobotan atribut ini kemudian digunakan sebagai masukan pada model DT dan Random Tree untuk membentuk struktur pohon keputusan yang lebih efisien dan akurat. Tabel 2 menunjukkan bobot atribut yang diperoleh setelah dilakukan proses optimasi menggunakan *PSO*.

Tabel 3. Bobot Atribut Model *Decision Tree* dan Random Tree setelah *PSO*

attribute	Bobot DT	Bobot Random Tree
smt	0.713	0
jen_kel	1	1
jrs_slta	0.844	0
klh_smbk_krj	0.779	0
pkj_ayah	0	0
sts_aktif_cuti	0.578	1

Dari Tabel 3 terlihat bahwa pada DT, atribut *jen_kel* memiliki bobot tertinggi (1,0) dan menjadi faktor dominan, sedangkan *jrs_slta*, *klh_smbk_krj*, dan *smt* juga berkontribusi penting. Atribut *pkj_ayah* memiliki bobot 0 dan diabaikan. Pada Random Tree, hanya *jen_kel* dan *sts_aktif_cuti* yang signifikan, sedangkan atribut lainnya diabaikan. Hal ini menunjukkan bahwa Random Tree lebih sederhana namun berpotensi kehilangan informasi penting, sesuai dengan performa klasifikasinya yang lebih rendah dibandingkan DT. Visualisasi *Feature Importance* pada Gambar 6 memperkuat interpretasi ini, menunjukkan bahwa *PSO* membantu menyoroti atribut yang paling relevan sekaligus menyederhanakan model.



Gambar 6. *Feature Importance*

Setelah dilakukan seleksi fitur menggunakan algoritma *PSO*, model klasifikasi kembali dibangun menggunakan algoritma DT dan Random Tree dengan confusion matrix terlihat pada tabel 4.

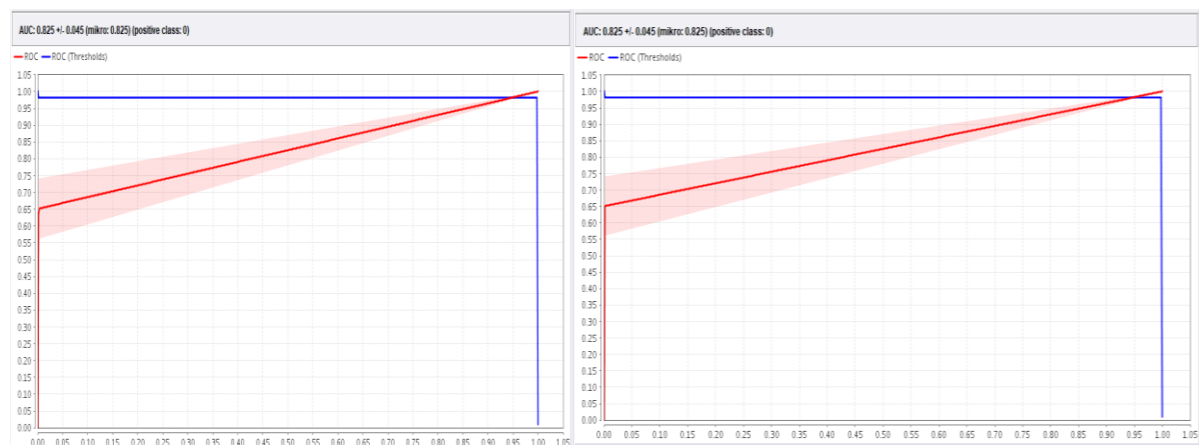
Tabel 4. Perbandingan Confusion Matrix

Model	TP	FP	FN	TN
DT	15.184	177	6	330
Random Tree	15.189	374	1	133
DT+ <i>PSO</i>	15.184	177	6	330
Random Tree+ <i>PSO</i>	15.184	177	6	330

Pada tabel 4, hasil yang diperoleh menunjukkan bahwa confusion matrix untuk *Decision Tree* dengan *PSO* identik dengan model *Decision Tree* tanpa seleksi fitur. Hasil akurasi model *Decision Tree* dengan optimasi *PSO* sebesar 98,83%, *precision* 98,20%, dan *recall* 65,13%. Meskipun tidak terjadi perubahan pada jumlah prediksi benar dan salah, terdapat sedikit perbedaan pada nilai *precision* dan *recall* dibandingkan model tanpa *PSO*. *Precision* menurun dari 98,21% menjadi 98,0% dan *Recall* meningkat dari 65,09% menjadi 65,13%. Perbedaan ini sangat kecil dan tidak signifikan secara praktis, namun menunjukkan bahwa seleksi fitur dengan *PSO* tidak menurunkan kinerja model dan justru mempertahankan stabilitas performa klasifikasi.

Confusion matrix dari model klasifikasi Random Tree yang telah dioptimasi dengan seleksi fitur menggunakan *PSO*. Menariknya, hasil ini identik dengan hasil model *Decision Tree* baik dengan maupun tanpa *PSO*, dan sangat berbeda dengan hasil Random Tree sebelum optimasi. Setelah dilakukan seleksi fitur dengan *PSO*, model Random Tree menghasilkan akurasi 98,83%, *precision* 98,23%, dan *recall* 65,10% sama persis dengan model *Decision Tree* yang telah dioptimasi.

Setelah dilakukan proses klasifikasi ulang dengan subset fitur hasil seleksi menggunakan *PSO*, kedua model, yakni *Decision Tree* dan Random Tree, dievaluasi menggunakan kurva ROC. Perbandingan kurva ROC dari kedua model setelah dilakukan seleksi fitur dengan *PSO*. Nilai Area Under Curve (*AUC*) digunakan sebagai indikator kuantitatif dari akurasi klasifikasi model secara keseluruhan terdapat pada Gambar 8.



Gambar 7. Kurva ROC *Decision* & *Random Tree* dengan *PSO*

Berdasarkan Gambar 7, baik model *Decision Tree* maupun Random Tree yang telah dioptimasi dengan *PSO* menunjukkan performa klasifikasi yang sangat baik, ditandai dengan nilai *AUC* sebesar 0.825 ± 0.045 . Kurva ROC untuk kedua model berada jauh di atas garis diagonal, menandakan kemampuan klasifikasi yang kuat dalam membedakan mahasiswa yang menunggak dan tidak. Peningkatan yang signifikan terlihat khususnya pada model Random Tree, yang sebelumnya hanya menghasilkan *AUC* sebesar 0.632 sebelum dilakukan seleksi fitur. Hal ini menunjukkan bahwa proses optimasi fitur dengan *PSO* tidak hanya menyederhanakan jumlah atribut yang digunakan, tetapi juga secara nyata meningkatkan akurasi dan kestabilan model.

Fakta bahwa kedua model menghasilkan nilai *AUC* yang identik menunjukkan bahwa dengan subset fitur yang optimal, performa klasifikasi antara *Decision Tree* dan Random Tree dapat disejajarkan. Ini membuktikan bahwa pemilihan atribut yang tepat memiliki dampak yang lebih signifikan terhadap hasil prediksi dibandingkan pemilihan algoritma semata.

3.5 Evaluasi Model

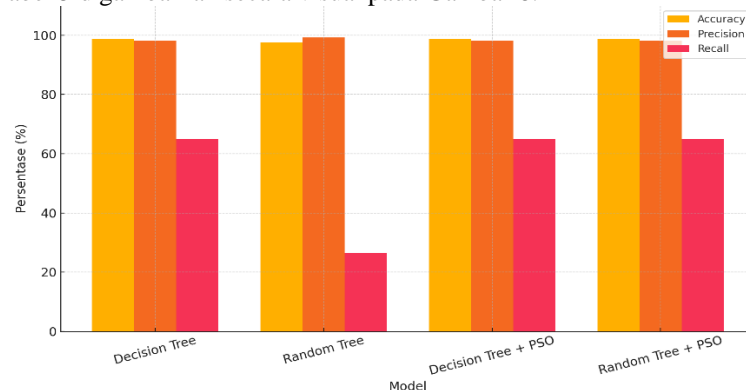
Evaluasi dilakukan terhadap empat skenario model, yaitu: *Decision Tree* tanpa *PSO*, Random Tree tanpa *PSO*, *Decision Tree* dengan *PSO*, dan Random Tree dengan *PSO*. Setiap model dievaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *AUC* untuk mengukur performa klasifikasi secara menyeluruh terlihat pada tabel 5.

Tabel 5. Perbandingan Hasil

Model	Akurasi	Presisi	<i>Recall</i>	F1	
				Score	<i>AUC</i>
DT	98,83%	98,21%	65,09%	82,50%	82,50%
Random Tree	97,61%	99,25%	26,33%	70,15%	63,20%
DT+ <i>PSO</i>	98,83%	98,20%	65,13%	82,50%	82,50%
Random Tree+ <i>PSO</i>	98,83%	98,23%	65,10%	82,50%	82,50%

Berdasarkan Tabel 5, model *Decision Tree* tanpa *PSO* memiliki akurasi tertinggi (98,83%) dibandingkan Random Tree tanpa *PSO* yang hanya mencapai 97,61%. Hal ini menunjukkan bahwa *Decision Tree* lebih andal dalam mengidentifikasi mahasiswa yang menunggak pembayaran secara tepat. Namun, Random Tree memiliki nilai *precision* lebih tinggi (99,25%), yang berarti lebih baik dalam meminimalisir kesalahan prediksi positif palsu (False Positive). Setelah diterapkan proses seleksi fitur menggunakan *PSO*, terjadi peningkatan signifikan pada performa model Random Tree. *AUC* Random Tree meningkat dari 0.632 menjadi 0.825, setara dengan *Decision Tree* + *PSO*. Hal ini membuktikan bahwa seleksi fitur berkontribusi besar terhadap peningkatan kemampuan klasifikasi model Random Tree, yang sebelumnya kurang optimal. Sementara pada model *Decision Tree*, performanya relatif stabil sebelum dan sesudah seleksi fitur, dengan akurasi tetap di angka 98,83% dan nilai *AUC* tetap tinggi (0.825 ± 0.045). Artinya, *PSO* tidak menurunkan kualitas model meskipun jumlah fitur berkurang. Menariknya, setelah dilakukan optimasi fitur dengan *PSO*, kedua model (*Decision Tree* dan Random Tree) menghasilkan confusion matrix dan metrik performa yang identik, yaitu: Akurasi: 98,83%, *Precision*: 98,23%, *Recall*: 65,10%, *AUC*: 0.825 ± 0.045 .

Hasil evaluasi menunjukkan bahwa meskipun nilai akurasi dan *precision* cukup tinggi, *recall* yang diperoleh relatif lebih rendah. Kondisi ini dapat dijelaskan oleh distribusi kelas pada dataset, yaitu mahasiswa dengan status tidak menunggak berjumlah 15.190, sedangkan mahasiswa yang menunggak hanya 507. Ketidakseimbangan kelas (imbalanced class) ini menyebabkan model cenderung lebih baik dalam mengenali kelas mayoritas (tidak menunggak), tetapi kurang sensitif terhadap kelas minoritas (menunggak). Akibatnya, masih ada sebagian mahasiswa yang menunggak tidak terdeteksi dengan baik, sehingga *recall* menjadi rendah. Fenomena ini umum terjadi pada permasalahan dengan distribusi kelas yang tidak seimbang. Untuk mempermudah interpretasi metrik performa, data dari Tabel 5 digambarkan secara visual pada Gambar 8.



Gambar 8. Diagram Batang Perbandingan Hasil Evaluasi Model

Berdasarkan Tabel 5 dan Gambar 8, terlihat bahwa *Decision Tree* tanpa *PSO* memiliki akurasi tertinggi (98,83%), sedangkan Random Tree tanpa *PSO* lebih rendah (97,61%). Namun, Random Tree memiliki nilai *precision* lebih tinggi (99,25%), menandakan model ini lebih efektif meminimalkan kesalahan prediksi positif palsu (False Positive). Setelah diterapkan seleksi fitur dengan *PSO*, performa Random Tree meningkat signifikan sehingga akurasi, *precision*, *recall*, *f1-score* dan *AUC* menjadi identik dengan *Decision Tree* yang telah dioptimasi, yaitu: Akurasi 98,83%, *Precision* 98,23%, *Recall* 65,10%, dan *AUC* 0.825 ± 0.045 . Hal ini menunjukkan bahwa seleksi fitur berkontribusi lebih besar terhadap peningkatan performa dibandingkan sekadar pemilihan algoritma.

Hasil pembobotan atribut menunjukkan bahwa jenis kelamin (*jen_kel*) menjadi faktor dominan, dengan bobot tertinggi pada DT maupun Random Tree setelah *PSO*. Secara statistik, *PSO* menilai kontribusi tiap fitur terhadap peningkatan akurasi klasifikasi, sehingga dominasi *jen_kel* menunjukkan korelasi yang konsisten dengan status mahasiswa aktif/cuti. Dari perspektif sosial, hal ini dapat mencerminkan perbedaan perilaku akademik atau kepatuhan administrasi antara mahasiswa laki-laki dan perempuan, sesuai dengan literatur sebelumnya, misalnya penelitian SVM+*PSO* [19] yang menunjukkan peningkatan akurasi dari 84,00% menjadi 98,85% setelah optimasi fitur, meskipun tidak menekankan kontribusi atribut spesifik.

Sebaliknya, atribut pekerjaan ayah (*pkj_ayah*) tidak berpengaruh, dengan bobot nol setelah *PSO*. Hal ini menunjukkan keterkaitan yang lemah antara pekerjaan ayah dan status pembayaran mahasiswa, atau homogenitas sampel sehingga atribut ini tidak menjadi pembeda signifikan. Temuan ini konsisten dengan literatur yang menekankan bahwa faktor ekonomi keluarga sebaiknya diukur dengan indikator lebih langsung, misalnya pendapatan, beasiswa, atau tanggungan keluarga, agar relevan terhadap outcome pembayaran.

Secara kebijakan, hasil ini memberikan beberapa implikasi: 1) Kampus dapat memfokuskan intervensi pada atribut yang relevan, seperti gender dan status aktif/cuti sebelumnya, untuk program pengingat pembayaran atau beasiswa. 2) Penyederhanaan model melalui *PSO* mengurangi kompleksitas administrasi data tanpa menurunkan performa prediksi. 3) Pendekatan ini lebih efisien, interpretable, dan aplikatif dibandingkan penelitian sebelumnya yang menggunakan SVM atau Random Forest tanpa optimasi fitur, karena memberikan kombinasi antara akurasi tinggi dan kemudahan interpretasi oleh pihak manajemen.

Meskipun demikian, terdapat beberapa keterbatasan dan potensi bias: 1) Dataset hanya mencakup variabel tertentu, sehingga faktor ekonomi lain yang berpengaruh mungkin tidak terwakili. 2) Proporsi mahasiswa laki-laki dan perempuan yang tidak seimbang dapat memengaruhi bobot *jen_kel*. 3) Hasil klasifikasi tergantung pada kualitas data numerik yang dihasilkan dari encoding, sehingga kesalahan pra-pemrosesan dapat memengaruhi performa. 4) dapat dipertimbangkan penerapan teknik penyeimbangan data seperti SMOTE atau undersampling untuk meningkatkan performa pada kelas minoritas.

Dengan mempertimbangkan hal-hal di atas, hasil penelitian menunjukkan bahwa integrasi algoritma pohon keputusan dengan seleksi fitur berbasis *PSO* tidak hanya mempertahankan akurasi dan stabilitas model, tetapi juga menyediakan dasar yang jelas untuk pengambilan keputusan manajemen kampus dalam mitigasi risiko keterlambatan pembayaran mahasiswa.

4. KESIMPULAN

Penelitian ini berhasil membangun model prediktif keterlambatan pembayaran mahasiswa menggunakan *Decision Tree* dan *Random Tree* dengan seleksi fitur berbasis *PSO*. Hasil menunjukkan bahwa seleksi fitur mampu menyederhanakan atribut tanpa menurunkan akurasi, serta meningkatkan performa model *Random Tree* dari *AUC* 0,632 menjadi 0,825. Temuan ini menegaskan bahwa pemilihan fitur lebih berpengaruh terhadap performa klasifikasi dibandingkan pemilihan algoritma. Secara praktis, model ini dapat diintegrasikan dengan sistem informasi akademik kampus untuk memberikan peringatan dini kepada mahasiswa berisiko menunggak pembayaran, sehingga manajemen dapat melakukan intervensi preventif secara tepat waktu.

REFERENCES

- [1] Ridwansyah, M. Iqbal, H. Destiana, Sugiono, and A. Hamid, "Data Mining Berbasis Machine Learning Untuk Analitik Prediktif Dalam Kelulusan," *semanTIK*, vol. 10, no. 2, pp. 1–10, 2024, doi: <https://doi.org/10.55679/semantik.v10i2.67>.
- [2] W. Li, "Design of Financial Crisis Early Warning Model Based on *PSO*-SVM Algorithm," *Math. Probl. Eng.*, pp. 1–8, 2022, doi: <https://doi.org/10.1155/2022/3241802>.
- [3] N. W. D. Ayuni, N. N. Lasmini, and K. C. Dewi, "Predicting financial distress of property and real estate companies using optimized support vector machine-particle swarm optimization (SVM-*PSO*)," *Bull. Soc. Informatics Theory Appl.*, vol. 8, no. 1, pp. 97–106, 2024.
- [4] S. Anam, M. R. A. Putra, Z. Fitriah, I. Yanti, N. Hidayat, and D. M. Mahanani, "Health Claim Insurance Prediction Using Support Vector Machine With Particle Swarm Optimization," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 2, pp. 0797–0806, 2023, doi: [10.30598/barekengvol17iss2pp0797-0806](https://doi.org/10.30598/barekengvol17iss2pp0797-0806).
- [5] A. H. Kahfi, T. Prihatin, Yudhistira, A. Sudradjat, and G. Wijaya, "THE RIGHT STEPS TOWARDS GRADUATION: NB-*PSO* SMART COMBINATION FOR STUDENT GRADUATION PREDICTION," *J. Tek. Inform.*, vol. 5, no. 2, pp. 607–614, 2024, doi: <https://doi.org/10.52436/1.jutif.2024.5.2.1889>.
- [6] Sumarna, I. Nawawi, Suhardjono, Hari Sugiarto, and D. Yuliandari, "MENINGKATKAN AKURASI PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN METODE ALGORITMA GENETIKA," *J. Inform. Manaj. dan Komput.*, vol. 16, no. 2, 2024, doi: <http://dx.doi.org/10.36723/juri.v16i2.706>.
- [7] W. Widayani and H. Harlana, "Analisis Support Vector Machine Untuk Pemberian Rekomendasi Penundaan Biaya Kuliah Mahasiswa," *J. Sains dan Inform.*, vol. 7, no. 1, pp. 20–27, 2021, doi: [10.34128/jsi.v7i1.268](https://doi.org/10.34128/jsi.v7i1.268).
- [8] N. Y. L. Gaol, "Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam *Decision Tree* dan Algoritma C4.5," *J. Inf. Teknol.*, vol. 2, pp. 23–29, 2020, doi: [10.37034/jidt.v2i1.22](https://doi.org/10.37034/jidt.v2i1.22).
- [9] H. Nurdin, I. Carolina, R. L. Andharsaputri, A. Wuryanto, and Ridwansyah, "Forward Selection as a Feature Selection Method in the SVM Kernel for Student Graduation Data," *Sink. J. dan Penelit. Tek. Inform.*, vol. 8, no. October, pp. 2531–2537, 2024, doi: [10.33395/sinkron.v8i4.14172](https://doi.org/10.33395/sinkron.v8i4.14172).
- [10] M. J. Budiman and Fanny Jouke Doringin, "PENERAPAN ALGORITMA C5.0 DALAM MEMPREDIKSI KETERLAMBATAN PEMBAYARAN BIAYA KULIAH DI UNKRISWINA SUMBA," *J. Ilmu Komput. Revolutioner*, vol. 8, no. 6, 2024.
- [11] Max Bramer, *Principles of Data Mining*. Springer London, 2020.
- [12] A. Hamid and Ridwansyah, "Optimizing Heart Failure Detection : A Comparison between Naïve Bayes and Particle Swarm Optimization," *Paradigma*, vol. 26, no. 1, pp. 30–36, 2024, doi: <https://doi.org/10.31294/p.v26i1.3284>.
- [13] T. Azhima, Y. Siswa, and W. J. Pranoto, "Implementasi Seleksi Fitur Information Gain Ratio Pada Algoritma Random Forest Untuk Model Data Klasifikasi Pembayaran Kuliah," *Din. Inform.*, vol. 15, no. 1, pp. 41–49, 2023.
- [14] M. Norhalimi and T. A. Y. Siswa, "Optimasi Seleksi Fitur Information Gain pada Algoritma Naïve Bayes dan K-Nearest Neighbor," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 3, pp. 237–255, 2022, doi: [10.14421/jiska.2022.7.3.237-255](https://doi.org/10.14421/jiska.2022.7.3.237-255).
- [15] M. Mubarakah, D. Aditya Nugraha, and A. Yunus, "Penerapan Algoritma C4.5 dalam Memprediksi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan," *J. Ris. Mhs. Bid. Teknol. Inf.*, vol. 7, no. 1, pp. 1–5, 2024, [Online]. Available: <https://ejournal.unikama.ac.id/index.php/JFTI>.
- [16] H. Umar, R. Kusumawati, M. Imamudin, and M. A. Rohman, "Klasifikasi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan Menggunakan Algoritma Naïve Bayes dan Support Vector Machine," *TIN Terap. Inform. Nusantara*, vol. 4, no. 11, pp. 709–718, 2024, doi: [10.47065/tin.v4i11.4969](https://doi.org/10.47065/tin.v4i11.4969).
- [17] T. A. Y. Siswa and R. P. Wibowo, "Komparasi Metode Seleksi Fitur Dalam Prediksi Keterlambatan Pembayaran Biaya Kuliah," *Teknika*, vol. 12, no. 1, pp. 73–82, 2023, doi: [10.34148/teknika.v12i1.601](https://doi.org/10.34148/teknika.v12i1.601).



- [18] M. R. Akhmad and T. A. Y. Siswa, “Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi,” *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 185, 2022, doi: 10.35889/progresif.v18i2.921.
- [19] H. Nurdin, I. Nawawi, A. Wuryanto, D. Yuliandari, and H. Sugiarto, “Prediksi Keterlambatan Pembayaran Mahasiswa untuk Mitigasi Risiko Cuti Menggunakan SVM Optimasi PSO,” *J. Apl. Sains, Inf. , Elektron. dan Komput.*, vol. 7, no. 1, pp. 1–11, 2025, doi: <https://doi.org/10.26905/jasiek.v7i1.15483>.
- [20] S. Mulyono, *Manajemen Keuangan Perguruan Tinggi*. Jakarta: PT RajaGrafindo Persada, 2010.
- [21] D. Ratnasari, M. Sari, and A. Putri, “Analisis Pembiayaan dan Manajemen Finansial di Perguruan Tinggi Swasta,” *J. Manaj. Pendidik.*, vol. 12, no. 2, pp. 45–58, 2025.
- [22] S. Mirjalili, *Nature-Inspired Optimizers: Theories, Literature Reviews and Applications*. Springer, 2019.