

PREDIKSI STATUS AKADEMIK MAHASISWA BERDASARKAN DATA PEMBAYARAN DENGAN NAIVE BAYES DAN PARTICLE SWARM OPTIMIZATION

Sinta Rukiastiandari¹, Luthfia Rohimah², Fara Mutia³, Aprillia⁴, Chodidjah⁵

¹Program Studi Teknologi Informasi, Universitas Bina Sarana Informatika, ²Program Studi Sistem Informasi PSDKU Bogor, Universitas Bina Sarana Informatika, ^{3,4,5}Program Studi Manajemen, Universitas Bina Sarana Informatika

^{1,2,3,4,5}Jln. Kramat Raya No.98, Senen, Jakarta Pusat 10450

¹sinta.sru@bsi.ac.id, ²luthfia.lhm@bsi.ac.id, ³fara.fru@bsi.ac.id, ⁴aprillia.prl@bsi.ac.id,

⁵chodidjah.cdj@bsi.ac.id

Abstract

Higher education in Indonesia faces challenges in managing student payments, where delays can impact academic status, including the risk of leave or withdrawal. This study aims to develop a predictive model for academic status based on tuition payment data using the Naive Bayes (NB) method optimized using Particle Swarm Optimization (PSO). The dataset consists of 15,697 student records that have undergone preprocessing, including handling missing values and encoding categorical attributes. The results show that the NB model produces an accuracy of 98.83%, a precision of 98.21%, a recall of 65.09%, and an AUC of 0.905. Optimization with PSO increases the recall to 65.13% and the AUC to 0.907, while accuracy and precision remain stable. Feature analysis indicates that Gender, High School Major, and Studying While Working are the most influential attributes, while Father's Occupation is relatively less significant. These findings confirm the potential of NB-PSO as a predictive approach to support more effective academic administration management.

Keywords : *Academic Status Prediction, Late Payment, Machine Learning, Naive Bayes, Particle Swarm Optimization.*

Abstrak

Pendidikan tinggi di Indonesia menghadapi tantangan dalam pengelolaan pembayaran mahasiswa, di mana keterlambatan dapat berdampak pada status akademik, termasuk risiko cuti atau pengunduran diri. Penelitian ini bertujuan mengembangkan model prediksi status akademik berbasis data pembayaran kuliah dengan metode Naive Bayes (NB) yang dioptimasi menggunakan Particle Swarm Optimization (PSO). Dataset berjumlah 15.697 record mahasiswa yang telah melalui pra-pemrosesan, termasuk penanganan nilai hilang dan pengkodean atribut kategorikal. Hasil menunjukkan bahwa model NB menghasilkan akurasi 98,83%, precision 98,21%, recall 65,09%, dan AUC 0,905. Optimasi dengan PSO meningkatkan recall menjadi 65,13% dan AUC menjadi 0,907, sementara akurasi dan precision tetap stabil. Analisis fitur mengindikasikan bahwa Jenis Kelamin, Jurusan SLTA, dan Kuliah Sambil Bekerja merupakan atribut paling berpengaruh, sedangkan Pekerjaan Ayah relatif kurang signifikan. Temuan ini menegaskan potensi NB-PSO sebagai pendekatan prediktif untuk mendukung pengelolaan administrasi akademik yang lebih efektif.

Kata kunci : *Keterlambatan Pembayaran, Machine Learning, Naive Bayes, Particle Swarm Optimization, Prediksi Status Akademik.*

1. PENDAHULUAN

Pendidikan tinggi di Indonesia menghadapi tantangan dalam pengelolaan administrasi keuangan mahasiswa, khususnya terkait

ketepatan pembayaran biaya pendidikan yang berpengaruh pada kelancaran studi [1]. Keterlambatan pembayaran dapat menimbulkan risiko administratif, seperti perubahan status

menjadi cuti atau pengunduran diri [2], serta berdampak pada menurunnya performa akademik mahasiswa [4].

Prediksi keterlambatan pembayaran menjadi penting karena dapat membantu kampus mengantisipasi masalah akademik sejak dini [3]. Selain itu, klasifikasi pembayaran tepat waktu maupun terlambat juga berperan sebagai indikator penting dalam manajemen administrasi pendidikan tinggi [5]. Namun, deteksi dini masih sering dilakukan secara manual sehingga kurang efektif [6]. Penerapan data mining dapat menjadi solusi untuk memprediksi status mahasiswa secara lebih akurat dan mendukung upaya mitigasi risiko cuti [7].

Sejumlah penelitian terdahulu membahas penerapan metode data mining dan machine learning dalam konteks akademik maupun finansial. Model prediksi kelulusan mahasiswa berbasis machine learning telah berhasil dikembangkan [8]. Model deteksi krisis keuangan juga telah dikaji menggunakan kombinasi algoritma Support Vector Machine dan Particle Swarm Optimization [9]. Pendekatan serupa diterapkan untuk memprediksi kesulitan keuangan pada perusahaan properti dan real estate [10], serta prediksi klaim asuransi kesehatan [11].

Dalam konteks pendidikan, kombinasi Naive Bayes (NB) dengan (PSO) Particle Swarm Optimization telah digunakan untuk memprediksi kelulusan mahasiswa [12]. Pendekatan lain juga diterapkan melalui seleksi fitur pada algoritma Random Forest untuk memodelkan keterlambatan pembayaran kuliah [13]. Metode seleksi fitur dalam prediksi keterlambatan pembayaran biaya kuliah juga telah dibandingkan [14], sementara algoritma K-Nearest Neighbor digunakan untuk kasus serupa [15].

Meskipun demikian, kajian yang secara khusus meneliti prediksi status akademik mahasiswa (aktif, cuti, atau mengundurkan diri) berdasarkan data pembayaran kuliah dengan pendekatan NB yang dioptimasi PSO masih jarang dilakukan [16]. Penelitian terdahulu yang menggunakan SVM-PSO [7], tetapi lebih berfokus pada mitigasi risiko cuti, bukan pada optimalisasi prediksi berbasis NB. Inilah celah penelitian (research gap) yang ingin dijawab melalui studi ini.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi status akademik mahasiswa berdasarkan data pembayaran kuliah menggunakan algoritma NB yang dioptimasi dengan PSO.

Kontribusi penelitian ini adalah:

1. Menyediakan sistem prediksi keterlambatan pembayaran mahasiswa yang dapat membantu pihak kampus dalam melakukan deteksi dini risiko cuti maupun pengunduran diri.

2. Membuktikan efektivitas kombinasi NB dengan PSO dalam meningkatkan akurasi prediksi pada domain pendidikan tinggi.

3. Menawarkan alternatif pendekatan baru yang berbeda dengan penelitian terdahulu yang lebih banyak berfokus pada algoritma Decision Tree, Random Forest, maupun SVM.

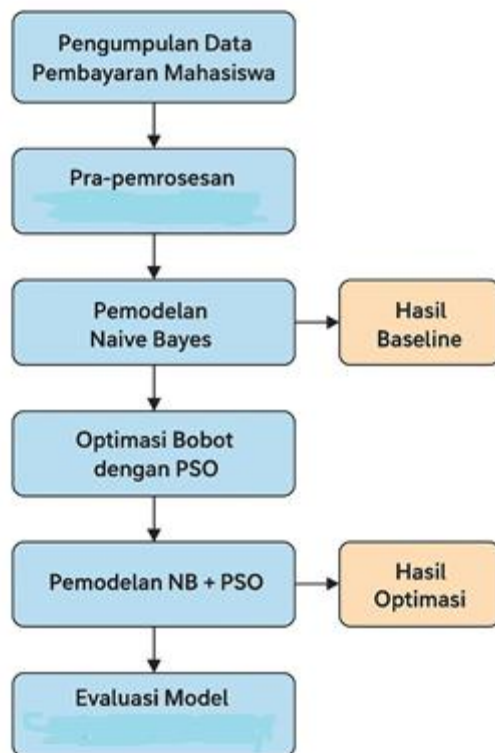
Kebaruan penelitian ini terletak pada penerapan NB-PSO secara khusus untuk memprediksi status akademik mahasiswa berdasarkan data pembayaran kuliah. Fokus ini membedakan penelitian ini dari studi sebelumnya yang umumnya menyoroti prediksi kelulusan atau analisis finansial di luar konteks administrasi mahasiswa.

2. METODOLOGI PENELITIAN

2.1. Skema Alur Penelitian

Secara garis besar, penelitian ini dilakukan melalui beberapa tahapan utama. Proses dimulai dari pengumpulan dan persiapan data pembayaran mahasiswa. Selanjutnya dilakukan tahap pra-pemrosesan berupa pembersihan data, termasuk penanganan nilai yang hilang. Setelah data siap, dilakukan dua skenario pemodelan, yaitu: 1). Model Baseline menggunakan algoritma Naive Bayes tanpa optimasi. 2). Model Optimasi dengan Naive Bayes yang ditingkatkan melalui Particle Swarm Optimization (PSO) menggunakan operator Optimize Weights di RapidMiner.

Hasil dari kedua model dievaluasi menggunakan skema Nominal X-Validation dengan metrik akurasi, precision, recall, dan AUC. Perbandingan hasil evaluasi ini kemudian dianalisis untuk menentukan efektivitas penggunaan PSO dalam meningkatkan performa Naive Bayes pada prediksi status akademik mahasiswa. Skema tersebut digambarkan dalam alur penelitian pada Gambar 1.



Gambar 1. Alur Penelitian

2.2. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan data pembayaran mahasiswa dengan total 15.697 record. Awalnya jumlah data lebih besar, namun sebagian record yang tidak lengkap telah dibuang secara manual sehingga hanya data yang valid dan bersih yang dipakai dalam pemodelan. Data ini mencakup informasi status pembayaran mahasiswa, baik yang tepat waktu maupun yang terlambat, yang selanjutnya digunakan untuk prediksi status akademik.

2.3. Data Preprocessing

Pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum digunakan dalam pemodelan [17]. Langkah-langkah yang dilakukan antara lain:

1. Pembersihan data secara manual dengan menghapus data yang tidak lengkap.
2. Import dengan menggunakan operator Read Excel pada RapidMiner.
3. Penanganan data hilang dengan menggunakan operator Replace Missing.

Hasil pra-pemrosesan ini menghasilkan dataset bersih dengan 15.697 record siap pakai untuk proses klasifikasi.

2.4. Penerapan Algoritma Naïve Bayes

Tahap awal pemodelan dilakukan menggunakan algoritma NB tanpa optimasi dengan skema Nominal X-Validation di RapidMiner sebagai baseline untuk dibandingkan dengan model setelah optimasi [18]. NB merupakan algoritma klasifikasi berbasis probabilistik yang menentukan kelas data berdasarkan distribusi nilai atribut dengan asumsi independensi antaratribut [19], [20]. Meskipun asumsi tersebut tidak selalu sesuai dengan data nyata, NB tetap efektif karena sederhana, cepat, dan cukup handal, sehingga banyak digunakan sebagai model dasar dalam penelitian klasifikasi, termasuk pada prediksi keterlambatan pembayaran mahasiswa [21].

2.5. Optimasi dengan PSO

Setelah memperoleh baseline, dilakukan optimasi menggunakan Particle Swarm Optimization (PSO) melalui operator Optimize Weights (PSO) di RapidMiner. Tahapannya meliputi: (1) import data dengan Read Excel, (2) penanganan data hilang menggunakan Replace Missing, (3) optimasi bobot fitur dengan Optimize Weights (PSO) untuk meminimalkan atribut kurang relevan, dan (4) klasifikasi hasil optimasi dengan Naive Bayes (NB) melalui Nominal X-Validation.

Parameter PSO ditetapkan dengan swarm size 5 partikel, max iterations 30, inertia weight (w) 1.0 (rentang 0.0–1.0), serta cognitive (c1) dan social coefficient (c2) masing-masing 1.0. Pengaturan ini bertujuan menyeimbangkan eksplorasi solusi baru dan eksploitasi solusi terbaik, sehingga diperoleh bobot fitur yang lebih optimal. Hasil optimasi kemudian diintegrasikan dengan NB untuk menghasilkan model dengan performa lebih baik dibandingkan baseline.

2.6. Evaluasi Model

Evaluasi dilakukan dengan membandingkan hasil model NB tanpa optimasi dan NB dengan optimasi PSO. Indikator evaluasi yang digunakan antara lain akurasi, precision, recall, dan AUC/ROC. Perbandingan ini bertujuan untuk melihat sejauh mana PSO mampu meningkatkan performa algoritma NB dalam memprediksi status akademik mahasiswa berdasarkan data pembayaran. Indikator evaluasi yang digunakan dengan rumus sebagai berikut:

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

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

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

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan dari kedua metode tersebut berdasarkan alur penelitian sebagai berikut.

3.1. Pengumpulan Data

Setelah tahap pra-pemrosesan, dataset final berjumlah 15.697 record. Tabel I menunjukkan atribut yang digunakan dalam penelitian beserta tipe dan fungsinya.

TABEL I. ATRIBUT PEMBAYARAN

Atribut	Tipe	Keterangan
Nomor Biaya Kuliah	Binomial	Label: "Sudah Bayar" atau "Belum Bayar"
Semester	Integer	Semester mahasiswa saat ini
Jenis Kelamin	Integer	0 = Laki-laki, 1 = Perempuan
Kuliah Sambil Bekerja	Integer	0 = Tidak, 1 = Ya
Pernah Cuti	Integer	0 = Tidak, 1 = Ya
Pekerjaan Ayah	Integer	Kode pekerjaan ayah mahasiswa

Jurusan SLTA	Integer	Kode jurusan pada saat lulus SLTA
--------------	---------	-----------------------------------

Pada Tabel I menjelaskan Atribut Nomor Biaya Kuliah digunakan sebagai label untuk klasifikasi, sedangkan atribut lainnya menjadi fitur input untuk model prediksi. Atribut seperti Semester, Jenis Kelamin, dan Kuliah Sambil Bekerja bersifat kategorikal atau diskret dan memiliki distribusi yang berbeda-beda. Misalnya, mayoritas mahasiswa tidak bekerja sambil kuliah, dan sebagian kecil pernah cuti. Atribut Pekerjaan Ayah dan Jurusan SLTA memberikan informasi latar belakang sosial-ekonomi dan pendidikan, yang dapat memengaruhi perilaku pembayaran mahasiswa.

Distribusi awal ini penting untuk memahami sebaran data sebelum dilakukan pemodelan klasifikasi, dan memastikan bahwa dataset sudah seimbang serta siap digunakan untuk evaluasi model.

3.2. Data Preprocessing

Pada tahap data preprocessing, dilakukan beberapa langkah untuk menyiapkan dataset sebelum digunakan dalam pemodelan. Langkah-langkah tersebut meliputi pembersihan data, penanganan data yang hilang, dan pengkodean atribut-atribut yang bersifat kategorikal. Struktur dataset yang telah diproses, dengan atribut-atribut yang telah siap untuk digunakan dalam tahap klasifikasi dilihat pada Gambar 2.

ExampleSet (15697 examples, 1 special attribute, 6 regular attributes)							
Row No.	no_by_klh	smt	jen_kel	jrs_slta	klh_smbkrj	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. Sample Struktur Dataset Setelah Data Preprocessing

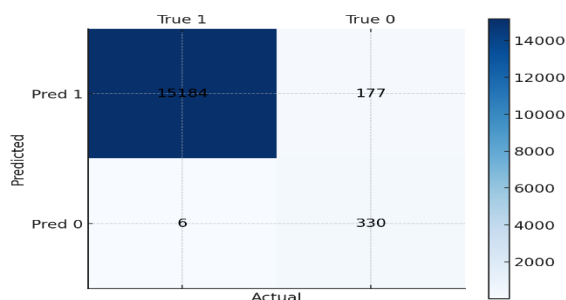
Gambar 2 ini menampilkan contoh dataset yang telah melalui tahap preprocessing, di mana atribut-atribut seperti Nomor Biaya Kuliah, Semester, Jenis Kelamin, dan lainnya telah diubah menjadi format yang sesuai untuk pemodelan.

Beberapa atribut dalam dataset, seperti Jenis Kelamin dan Kuliah Sambil Bekerja, merupakan data kategorikal yang perlu dikodekan menjadi format numerik agar dapat diproses oleh algoritma ID3. Misalnya, Jenis Kelamin yang

memiliki dua kategori yaitu "Laki-laki" dan "Perempuan", dikodekan menjadi 0 dan 1. Semester, yang dapat memiliki nilai dari 1 hingga beberapa semester. Dari Data yang hilang telah ditangani, dan semua atribut kategorikal telah dikodekan menjadi bentuk numerik yang dapat diproses oleh model NB.

3.3. Penerapan Algoritma NB

Untuk mengevaluasi performa model ID3 dalam mengklasifikasikan jenis layanan pembayaran digital, digunakan confusion matrix sebagai alat analisis. Confusion matrix memberikan gambaran detail mengenai prediksi model terhadap dua kelas: OVO dan GoPay, serta membandingkannya dengan label sebenarnya. *Confusion matrix* membantu dalam mengidentifikasi jumlah prediksi yang benar dan salah untuk masing-masing kategori. Hasil dari confusion matrix ditunjukkan pada Gambar 4.

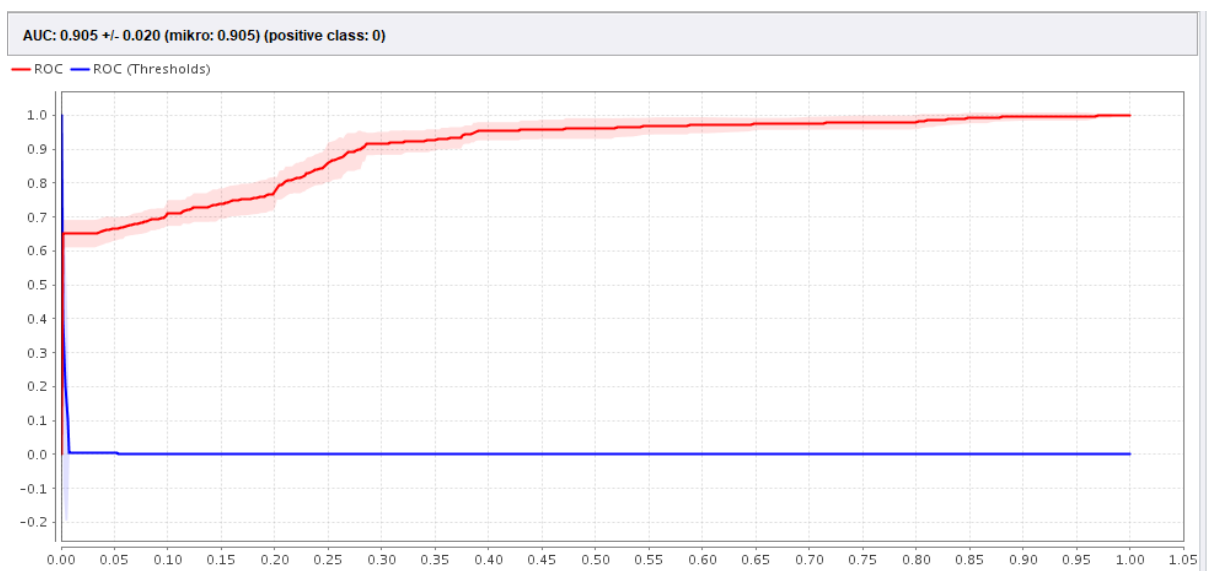


Gambar 3. *Confusion Matrix NB*

Pada Gambar 3, ditampilkan hasil *confusion matrix* dari prediksi keterlambatan pembayaran mahasiswa menggunakan metode NB. Dari matriks tersebut, diperoleh TP sebanyak 15.184 mahasiswa, yaitu mahasiswa yang diprediksi belum bayar dan memang benar faktanya. FP tercatat sebanyak 177 mahasiswa, yang diprediksi belum bayar tetapi faktanya sudah bayar. FN terdapat 6 mahasiswa, yang diprediksi sudah bayar padahal sebenarnya belum bayar. Sedangkan TN sebanyak 303 mahasiswa, yaitu mahasiswa yang diprediksi belum bayar dan memang benar faktanya.

Berdasarkan nilai-nilai tersebut, dapat dilakukan perhitungan metrik evaluasi kinerja model. Akurasi dihitung dengan membandingkan jumlah prediksi yang benar terhadap total data, *precision* menunjukkan kemampuan model dalam memprediksi mahasiswa yang benar-benar belum bayar, dan *recall* mengukur kemampuan model dalam mendeteksi seluruh mahasiswa yang belum bayar. Hasil perhitungan metrik evaluasi dari confusion matrix ini menghasilkan akurasi dengan rumus (1) sebesar 98.83%, *precision* dengan rumus (2) sebesar 98.21%, dan *recall* dengan rumus (3) sebesar 65.09%.

Hasil evaluasi tersebut kemudian diperkuat dengan analisis menggunakan kurva ROC dengan melihat nilai AUC pada ROC tersebut. Terlihat di Gambar 4.



Gambar 4. *Kurva ROC Model NB*

Berdasarkan Gambar 4, terlihat bahwa kurva ROC cenderung mendekati sisi kiri atas grafik. Hal ini mengindikasikan bahwa model Naïve Bayes

memiliki tingkat TP Rate yang tinggi dengan FP Rate yang relatif rendah pada sebagian besar titik ambang. Dengan kata lain, model mampu

membedakan mahasiswa yang belum bayar dan sudah bayar dengan cukup efektif.

Nilai AUC sebesar 0.905 ± 0.020 juga menegaskan bahwa performa klasifikasi model berada dalam kategori sangat baik (excellent). Menurut interpretasi standar, nilai AUC di atas 0.90 menandakan model memiliki keandalan tinggi dalam memisahkan kelas positif dan negatif. Walaupun masih terdapat beberapa kesalahan klasifikasi (FN maupun FP).

3.4. Optimasi dengan PSO

Setelah memperoleh hasil evaluasi awal menggunakan metode NB, penelitian ini kemudian dilanjutkan dengan melakukan proses optimasi menggunakan PSO dengan cara mencari bobot atribut yang lebih optimal sehingga mampu mengurangi kesalahan klasifikasi.

TABEL II. BOBOT ATRIBUT NB+PSO

Atribut	Bobot	Atribut	Bobot
semester	0.713	Pernah Cuti	0.578
Jenis kelamin	1	Pekerjaan Ayah	0
Kuliah Sambil bekerja	0.779	Jurusan SLTA	0.844

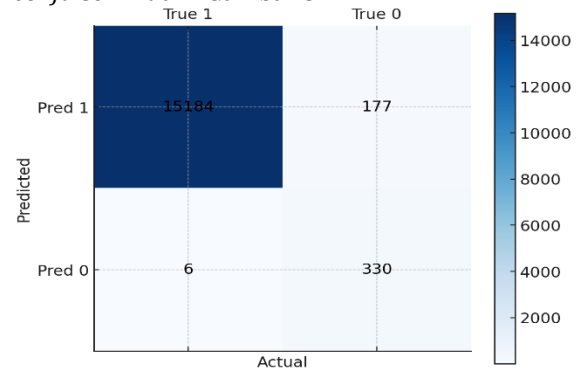
Tabel II merepresentasikan tingkat kepentingan (*importance*) masing-masing atribut dalam proses klasifikasi keterlambatan pembayaran mahasiswa. Dari tabel dapat dilihat bahwa: Jenis kelamin memiliki bobot tertinggi yaitu 1, artinya atribut ini dianggap paling berpengaruh dalam menentukan prediksi.

Jurusan SLTA juga memiliki bobot yang relatif tinggi sebesar 0.844, diikuti oleh Kuliah sambil bekerja dengan bobot 0.779. Hal ini menunjukkan bahwa latar belakang pendidikan serta status pekerjaan mahasiswa saat kuliah cukup relevan dalam memengaruhi ketepatan pembayaran.

Semester mendapat bobot 0.713 yang menandakan faktor akademik (tingkatan semester) juga berperan penting terhadap keterlambatan pembayaran. Pernah cuti memperoleh bobot 0.578, masih memberikan pengaruh meskipun tidak sebesar atribut lainnya.

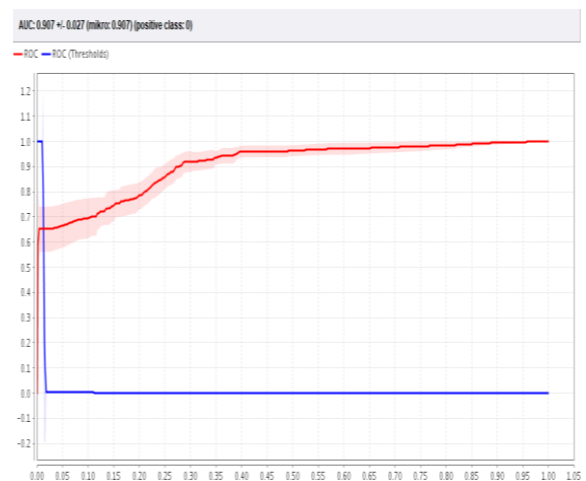
Sementara itu, pekerjaan ayah memiliki bobot 0.000, yang berarti atribut ini tidak memberikan kontribusi signifikan terhadap prediksi dan cenderung diabaikan oleh model setelah optimasi PSO.

Dari bobot tersebut akan menghasilkan *confusion matrix* Gambar 5.



Gambar 5. *Confusion Matrix NB+PSO*

Berdasarkan Gambar 5, diperoleh TP sebanyak 15.184 mahasiswa, yaitu mahasiswa yang diprediksi belum bayar dan memang benar belum membayar. FP sebanyak 177 mahasiswa, yaitu mahasiswa yang diprediksi belum bayar tetapi sebenarnya sudah membayar. FN tercatat sebanyak 6 mahasiswa, yakni mahasiswa yang diprediksi sudah membayar padahal faktanya belum membayar. Sementara itu, TN sebanyak 303 mahasiswa, yaitu mahasiswa yang diprediksi sudah membayar dan memang benar sudah membayar. Dari nilai tersebut, dilakukan perhitungan metrik evaluasi kinerja model dengan rumus (1), (2) dan (3). Hasil yang diperoleh yaitu akurasi 98.83%, precision 98.20%, dan recall 65,13%. Dari hasil tersebut akan menghasilkan kurva Gambar 6.



Gambar 6. *Kurva ROC Model NB+PSO*

Berdasarkan Gambar 6, terlihat bahwa kurva ROC pada model NB dan NB+PSO sama-sama mendekati sisi kiri atas grafik. Pencapaian nilai AUC sebesar 0.905 ± 0.027 menunjukkan bahwa

kemampuan model dalam membedakan kelas positif dan negatif berada pada tingkat yang sangat baik.

3.5. Evaluasi Model

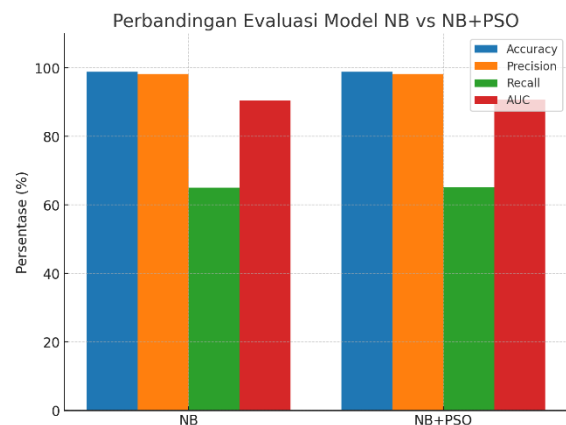
Evaluasi model dilakukan untuk membandingkan kinerja metode NB dan NB+PSO pada Tabel III.

TABEL III. PERBANDINGAN NB DAN NB+PSO

Metode	NB	NB+PSO
Accuracy	98,83%	98,83%
Precision	98,21%	98,20%
Recall	65,09%	65,13%
AUC	90,50%	90,70%

Tabel 3 menampilkan hasil evaluasi kinerja NB dan NB+PSO. Secara umum, kedua model menunjukkan performa yang hampir sebanding dengan akurasi yang sama, yaitu 98,83%. Hal ini menunjukkan bahwa Naive Bayes sudah cukup efektif sebagai baseline model karena karakteristik data yang relatif sederhana dan terstruktur dengan baik [18]. Perbedaan pada *precision* juga sangat kecil, sehingga dapat dikatakan tidak memberikan dampak signifikan.

Peningkatan terlihat pada metrik recall dan AUC, meskipun tipis, yaitu recall dari 65,09% menjadi 65,13% dan AUC dari 0,905 menjadi 0,907. Kenaikan ini sejalan dengan peran PSO yang lebih berfokus pada penyesuaian bobot fitur untuk meningkatkan kemampuan model dalam mengenali kelas minoritas. Namun, karena distribusi data cenderung imbalanced (jumlah mahasiswa tidak menunggak jauh lebih besar daripada yang menunggak), dampak optimasi tidak terlalu besar. Kondisi ini sejalan dengan temuan penelitian sebelumnya yang menyebutkan bahwa optimasi berbasis metaheuristik sering memberikan perbaikan terbatas pada model sederhana seperti NB, terutama ketika akurasi baseline sudah sangat tinggi. Perbandingan hasil evaluasi model ini divisualisasikan dalam bentuk diagram batang pada Gambar 7.



Gambar 7. Diagram Batang NB+PSO

Dari diagram tersebut terlihat bahwa meskipun peningkatan yang diperoleh tidak signifikan, penerapan PSO tetap menunjukkan kontribusi positif pada aspek penting, yaitu kemampuan model dalam mengenali mahasiswa yang benar-benar terlambat bayar (recall) serta meningkatkan diskriminasi kelas (AUC). Hal ini menegaskan bahwa optimasi berbasis PSO relevan diterapkan, terutama pada konteks prediksi di mana deteksi kasus minoritas lebih penting daripada peningkatan akurasi global.

4. Kesimpulan dan Saran

Model Naive Bayes (NB) tanpa optimasi telah mencapai akurasi 98,83%, precision 98,21%, recall 65,09%, dan AUC 0,905. Penerapan optimasi PSO memberikan peningkatan tipis pada recall (65,13%) dan AUC (0,907), sehingga terbukti mampu memperbaiki kemampuan deteksi mahasiswa yang benar-benar menunggak pembayaran. Hasil optimasi juga menunjukkan bahwa atribut Jenis Kelamin, Jurusan SLTA, dan Kuliah Sambil Bekerja memiliki kontribusi dominan, sedangkan atribut seperti Pekerjaan Ayah cenderung diabaikan.

Penelitian ini menegaskan bahwa kombinasi NB+PSO efektif sebagai metode prediksi keterlambatan pembayaran mahasiswa berbasis data akademik dan keuangan. Ke depan, penerapan optimasi bobot pada algoritma lain seperti K-NN, Neural Network, atau Gradient Boosting dapat memperluas pemahaman kontribusi fitur terhadap prediksi status akademik mahasiswa.

Daftar Pustaka:

- [1] 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>.
- [2] W. Widayani and H. Harliana, "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).
- [3] M. J. Budiman and Fanny Jouke Doringin, "PENERAPAN ALGORITMA C5.0 DALAM MEMPREDIKSI KETERLAMBATAN PEMBAYARAN BIAYA KULIAH DI UNKRISWINA SUMBA," *J. Ilmu Komput. Revolusioner*, vol. 8, no. 6, 2024.
- [4] M. Mubarokah, 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>.
- [5] 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).
- [6] 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).
- [7] 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>.
- [8] 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>.
- [9] 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>.
- [10] 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.
- [11] 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).
- [12] 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>.
- [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] 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).
- [15] 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.
- [16] A. Hamid and Ridwansyah, "Optimizing Heart Failure Detection : A Comparison between Naive Bayes and Particle Swarm Optimization," *Paradigma*, vol. 26, no. 1, pp. 30–36, 2024, doi: <https://doi.org/10.31294/p.v26i1.3284>.
- [17] Max Bramer, *Principles of Data Mining*. Springer London, 2020.
- [18] S. Mirjalili, *Nature-Inspired Optimizers: Theories, Literature Reviews and Applications*. Springer, 2019.
- [19] H. Nurdin, Suhardjono, A. Wuryanto, D. Yuliandari, and H. Sugiarto, "Naive Bayes and Particle Swarm Optimization in Early Detection Of Chronic Kidney Disease," *J. Tek. Inform.*, vol. 5, no. 3, pp. 703–708, 2024, doi: <https://doi.org/10.52436/1.jutif.2024.5.3>
- .1750.
- [20] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques*, (4th ed.). Morgan Kaufmann, 2022.
- [21] C. C. Aggarwal, *Data Mining: The Textbook*. Springer International Publishing, 2015.
- [22] V. Riyanto, H. Destiana, T. Prihatin, Sugiono, and G. Wijaya, "MENGOPTIMALKAN PREDIKSI GAGAL JANTUNG DENGAN KOMBINASI," *JIRE (Jurnal Inform. Rekayasa Elektron.*, vol. 8, no. 1, pp. 103–111, 2025, doi: <https://doi.org/10.36595/jire.v8i1.1541>.
- [23] S. Sartini, S. Sumarna, A. Hamid, A. H. Kahf, and Nicodias Palasar, "REVOLUSI DIAGNOSIS : OPTIMASI RANDOM TREE- PSO UNTUK PENYAKIT GINJAL KRONIS," *JIRE (Jurnal Inform. Rekayasa Elektron.*, vol. 8, no. 1, pp. 149–158, 2025, doi: <https://doi.org/10.36595/jire.v8i1.1542>.