

KOMPARASI ALGORITMA C4.5 (DECISION TREE) DAN NAIVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK KLASIFIKASI KELAYAKAN KREDIT DEBITUR

Gede Erik Aktama

Universitas Parna Raya

gedeerikaktama@gmail.com

Abstrak

Penyaluran pembiayaan kredit pada lembaga keuangan mikro dan koperasi simpan pinjam memiliki peranan strategis dalam mendorong pertumbuhan Usaha Mikro, Kecil, dan Menengah (UMKM). Namun demikian, tingginya risiko kredit bermasalah atau Non-Performing Loan (NPL) menjadi ancaman utama yang dapat mengganggu stabilitas likuiditas lembaga keuangan apabila proses analisis kelayakan calon debitur masih dilakukan secara subjektif dan manual. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma data mining klasifikasi terkemuka, yaitu C4.5 (Decision Tree) dan Naive Bayes, serta mengintegrasikan metode seleksi fitur berbasis algoritma metaheuristik Particle Swarm Optimization (PSO) guna meningkatkan akurasi prediksi kelayakan kredit. Dataset yang dianalisis mencakup 600 data riwayat nasabah pemohon kredit pada lembaga keuangan mikro di Sulawesi Utara dengan 12 variabel prediktor finansial dan demografis (antara lain: pendapatan bulanan, jumlah tanggungan, status kepemilikan agunan, riwayat kredit, rasio utang terhadap pendapatan / DTI, dan jangka waktu pinjaman). Evaluasi model dilakukan menggunakan teknik validasi 10-Fold Cross Validation untuk menghindari bias overfitting. Hasil eksperimen menunjukkan bahwa integrasi seleksi fitur PSO berhasil mereduksi 5 atribut yang bersifat redundan dan tidak relevan, menyisakan 7 fitur paling berpengaruh terhadap ketepatan prediksi. Penerapan PSO terbukti meningkatkan performa kedua algoritma secara signifikan: akurasi C4.5 meningkat dari 82,50% menjadi 89,17% (peningkatan +6,67%) dengan nilai Area Under ROC Curve (AUC) mencapai 0,918 (kategori Excellent Classification). Sementara itu, akurasi Naive Bayes meningkat dari 79,83% menjadi 84,67% (peningkatan +4,84%) dengan AUC sebesar 0,875. Model C4.5 berbasis PSO mencatatkan presisi sebesar 88,45%, recall sebesar 90,12%, dan F1-Score sebesar 89,28%, mengungguli Naive Bayes pada seluruh metrik evaluasi. Pohon keputusan yang dihasilkan oleh C4.5 juga memberikan kaidah klasifikasi (rule-based) yang transparan dan mudah diinterpretasikan oleh staf analis kredit dalam memitigasi risiko gagal bayar.

Kata kunci: Risiko Kredit, Klasifikasi Data Mining, Algoritma C4.5, Naive Bayes, Particle Swarm Optimization, Seleksi Fitur.

Abstract

Credit loan provisioning within microfinance institutions and credit unions serves as a critical catalyst for accelerating the expansion of Micro, Small, and Medium Enterprises (MSMEs). Nonetheless, elevated Non-Performing Loan (NPL) default rates pose severe systemic risks to institutional liquidity when creditworthiness appraisals rely upon subjective, manual assessments. This study aims to comparatively evaluate the predictive accuracy of two foundational machine learning classification algorithms—C4.5 (Decision Tree) and Naive Bayes—while integrating a metaheuristic Particle Swarm Optimization (PSO) feature selection paradigm to enhance scoring precision. The empirical dataset comprises 600 historical loan applicant records sourced from microfinance institutions in North Sulawesi, characterized across 12 demographic and financial attributes (including monthly net income, family dependents, collateral valuation, historical repayment track records, debt-to-income ratio, and loan tenor). Model robustness was verified via 10-Fold Cross-Validation. Experimental results demonstrate that PSO successfully filtered out 5 redundant attributes, retaining 7 optimal predictor features. Feature selection substantially improved both paradigms: C4.5 accuracy surged from 82.50% to 89.17% (+6.67% improvement) with an Area Under the ROC Curve (AUC) of 0.918 (classified as Excellent Classification), whereas Naive Bayes accuracy advanced from 79.83% to 84.67% (+4.84%) with an AUC of 0.875. The PSO-optimized C4.5 model achieved a precision of 88.45%, recall of 90.12%, and an F1-Score of 89.28%, decisively outperforming Naive Bayes across all evaluation metrics.

Furthermore, the explicit rule induction structure of C4.5 provides transparent, interpretable logic rules for credit underwriting officers to systematically mitigate default risks.

Keywords: Credit Risk Scoring, Data Mining Classification, C4.5 Algorithm, Naive Bayes, Particle Swarm Optimization, Feature Selection.

1. PENDAHULUAN

Sektor Usaha Mikro, Kecil, dan Menengah (UMKM) memegang peranan krusial sebagai tulang punggung perekonomian daerah di Sulawesi Utara. Dalam mendukung permodalan usaha tersebut, lembaga keuangan mikro, koperasi simpan pinjam, dan perbankan daerah secara aktif menyalurkan produk pinjaman kredit. Kendati demikian, risiko terjadinya kredit macet atau Non-Performing Loan (NPL) selalu menjadi ancaman terbesar bagi kesehatan likuiditas lembaga keuangan (Aktama et al., 2024; Kasmir, 2018). Kredit bermasalah yang melampaui ambang batas toleransi (standar Bank Indonesia maksimal 5%) dapat mengikis modal lembaga dan menghentikan siklus pembiayaan.

Pada praktiknya, sebagian besar lembaga keuangan mikro masih mengandalkan analisis kelayakan kredit secara konvensional berbasis prinsip 5C (Character, Capacity, Capital, Collateral, Condition). Meskipun prinsip 5C memiliki landasan konseptual yang baik, penilaiannya di lapangan sangat bergantung pada intuisi personal staf analis kredit (credit officer). Subjektivitas ini membuka celah timbulnya moral hazard, inkonsistensi penilaian, dan lambatnya proses persetujuan berkas pinjaman (Handayani & Hendrawan, 2020; Thomas et al., 2017). Oleh karena itu, penerapan teknik data mining dan machine learning untuk membangun model credit scoring otomatis menjadi instrumen krusial dalam mempercepat proses underwriting sekaligus memitigasi risiko gagal bayar.

Dalam ranah klasifikasi data mining, algoritma C4.5 (Decision Tree) yang dikembangkan oleh Quinlan (1993) dan algoritma probabilistik Naive Bayes (Domingos & Pazzani, 1997) merupakan dua algoritma yang paling banyak diterapkan. C4.5 memiliki keunggulan berupa kemampuannya menghasilkan pohon keputusan dan aturan logika (IF-THEN rules) yang mudah dipahami dan diaudit oleh praktisi perbankan. Di sisi lain, Naive Bayes dikenal sangat cepat dalam komputasi dan tangguh terhadap data yang tidak lengkap dengan memanfaatkan teorema probabilitas bersyarat Bayes (Han et al., 2022). Namun demikian, kedua algoritma ini rentan mengalami penurunan akurasi apabila dilatih pada dataset yang memuat banyak atribut yang tidak relevan (irrelevant attributes) atau atribut yang berkorelasi tinggi (redundant attributes) (Witten et al., 2016).

Untuk mengatasi permasalahan dimensionalitas data tersebut, teknik seleksi fitur (feature selection) berbasis algoritma metaheuristik Particle Swarm Optimization (PSO) menjadi pendekatan yang sangat menjanjikan (Kennedy & Eberhart, 1995). PSO mensimulasikan perilaku kawanan burung dalam mencari solusi ruang fitur optimal secara cepat dan adaptif. Integrasi PSO pada algoritma klasifikasi terbukti mampu mengeliminasi atribut pengganggu, mempercepat waktu pelatihan, dan mendongkrak performa akurasi secara signifikan (Xue et al., 2016).

Penelitian ini bertujuan untuk melakukan studi komparasi antara algoritma C4.5 dan Naive Bayes yang diintegrasikan dengan seleksi fitur Particle Swarm Optimization (PSO) dalam mengklasifikasikan kelayakan kredit debitur. Kontribusi penelitian ini mencakup: (1) identifikasi fitur-fitur penentu risiko kredit yang paling berpengaruh; (2) pembuktian empiris peningkatan akurasi C4.5 dan Naive Bayes setelah optimasi PSO; serta (3) formulasi aturan keputusan (decision rules) praktis bagi analis kredit.

2. METODOLOGI PENELITIAN

Metodologi penelitian mengacu pada tahapan Cross-Industry Standard Process for Data Mining (CRISP-DM) yang meliputi Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment (Chapman et al., 2000).

A. Dataset dan Variabel Penelitian

Penelitian menggunakan 600 sampel data historis nasabah pemohon kredit periode 2022-2023. Dataset terdiri dari 12 atribut prediktor dan 1 atribut label kelas (Layak / Tidak Layak):

- X1: Usia Debitur (Tahun)
- X2: Status Pernikahan (Belum / Menikah / Cerai)
- X3: Jumlah Tanggungan Keluarga (Orang)
- X4: Tingkat Pendidikan (SMA / Diploma / Sarjana)
- X5: Pendapatan Bersih Bulanan (Rupiah)
- X6: Status Pekerjaan / Usaha (Wiraswasta / PNS / Karyawan Swasta)
- X7: Lama Usaha / Bekerja (Tahun)
- X8: Nilai Taksiran Agunan / Jaminan (Rupiah)
- X9: Jumlah Pinjaman yang Diajukan (Rupiah)
- X10: Jangka Waktu Pinjaman / Tenor (Bulan)
- X11: Rasio Utang terhadap Pendapatan (Debt to Income Ratio / DTI)
- X12: Riwayat Riil Kredit Sebelumnya (Lancar / Macet)
- Y: Status Kelayakan Kredit (Layak [420 data] / Tidak Layak [180 data])

B. Algoritma Klasifikasi dan Optimasi PSO

1. Algoritma C4.5: Membentuk pohon keputusan dengan memilih atribut pemecah (split attribute) berdasarkan nilai Gain Ratio tertinggi:

$$\text{GainRatio}(S, A) = \frac{\text{Gain}(S, A)}{\text{SplitInfo}(S, A)}$$

di mana $\text{Gain}(S, A)$ adalah selisih antara entropi awal S dengan penjumlahan terbobot entropi subset atribut A .

2. Algoritma Naive Bayes: Menghitung probabilitas posterior $P(C|X)$ berdasarkan teorema Bayes dengan asumsi independensi fitur:

$$P(C_k | X) = \frac{P(C_k) * \prod_{i=1}^n P(x_i | C_k)}{P(X)}$$

3. Particle Swarm Optimization (PSO): Setiap partikel merepresentasikan subset biner fitur (1 = dipilih, 0 = dibuang). Posisi dan kecepatan partikel diperbarui berdasarkan pbest (posisi terbaik individu) dan gbest (posisi terbaik global kawanan):

$$v_i^{(t+1)} = w * v_i^t + c1 * r1 * (pbest_i - x_i^t) + c2 * r2 * (gbest - x_i^t)$$

$$x_i^{(t+1)} = \text{Sigmoid}(v_i^{(t+1)})$$

Fungsi fitness yang dimaksimalkan adalah nilai akurasi klasifikasi 10-Fold Cross-Validation.

3. HASIL PENELITIAN DAN PEMBAHASAN

Tahapan optimasi PSO dijalankan dengan parameter: ukuran populasi kawanan (swarm size) = 30 partikel, bobot inersia $w = 0,729$, konstanta akselerasi $c1 = c2 = 1,494$, dan iterasi maksimum = 50. Hasil seleksi fitur mereduksi dimensi data dari 12 atribut menjadi 7 atribut optimal, sebagaimana ditampilkan pada Tabel 1.

Tabel 1. Hasil Seleksi Fitur Menggunakan Particle Swarm Optimization (PSO)

Kode Fitur	Nama Atribut Data	Status Seleksi PSO	Bobot Pengaruh
------------	-------------------	--------------------	----------------

X1	Usia Debitur	Dieliminasi (0)	Rendah
X2	Status Pernikahan	Dieliminasi (0)	Rendah
X3	Jumlah Tanggungan Keluarga	Terpilih (1)	Sedang (0.642)
X4	Tingkat Pendidikan	Dieliminasi (0)	Rendah
X5	Pendapatan Bersih Bulanan	Terpilih (1)	Sangat Tinggi (0.912)
X6	Status Pekerjaan	Dieliminasi (0)	Rendah
X7	Lama Bekerja / Usaha	Terpilih (1)	Sedang (0.584)
X8	Nilai Agunan / Jaminan	Terpilih (1)	Tinggi (0.845)
X9	Jumlah Pinjaman	Terpilih (1)	Tinggi (0.795)
X10	Jangka Waktu Tenor	Dieliminasi (0)	Rendah
X11	Rasio Utang / DTI	Terpilih (1)	Sangat Tinggi (0.925)
X12	Riwayat Kredit Sebelumnya	Terpilih (1)	Sangat Tinggi (0.948)

Berdasarkan Tabel 1, tujuh atribut terpilih merefleksikan kombinasi kemampuan bayar (kapasitas finansial X5 dan rasio DTI X11), iktikad pembayaran (karakter X12), mitigasi risiko (agunan X8), serta beban pembiayaan (X3, X7, X9). Atribut demografis statis seperti usia, status pernikahan, dan tingkat pendidikan berhasil dieliminasi karena tidak memiliki korelasi yang signifikan terhadap risiko gagal bayar.

Tabel 2. Perbandingan Kinerja Algoritma Sebelum dan Sesudah Optimasi PSO

Model Klasifikasi	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)	AUC-ROC
Naive Bayes Standar	79.83%	78.40%	82.15%	80.23%	0.824
Naive Bayes + PSO	84.67%	83.90%	86.45%	85.15%	0.875
C4.5 (Decision Tree) Standar	82.50%	81.20%	84.30%	82.72%	0.862
C4.5 + PSO (Model Usulan)	89.17%	88.45%	90.12%	89.28%	0.918

Data pada Tabel 2 menunjukkan peningkatan performa yang sangat signifikan pada kedua model setelah diintegrasikan dengan seleksi fitur PSO. Algoritma Naive Bayes mengalami peningkatan akurasi sebesar +4,84% (dari 79,83% menjadi 84,67%). Peningkatan tertinggi dicatatkan oleh model C4.5 berbasis PSO yang mencapai akurasi 89,17% (naik +6,67% dari 82,50%). Nilai AUC C4.5 + PSO mencapai 0,918 yang tergolong dalam klasifikasi prima (Excellent Classification). Tingginya nilai recall (90,12%) sangat krusial dalam konteks pembiayaan, karena mencerminkan kemampuan model dalam meminimalkan debitur berisiko tinggi yang salah terprediksi sebagai debitur layak (False Positive minimization).

A. Kaidah Keputusan (Decision Rules) yang Terbentuk

Pohon keputusan yang dihasilkan oleh C4.5 berbasis PSO menghasilkan 8 aturan klasifikasi utama. Dua aturan paling representatif adalah:

- Rule 1: IF Riwayat_Kredit = 'Macet' AND Rasio_DTI > 40% THEN Status = 'Tidak Layak' (Confidence = 96,2%)
- Rule 2: IF Riwayat_Kredit = 'Lancar' AND Rasio_DTI <= 35% AND Nilai_Agunan >= 150%_Pinjaman THEN Status = 'Layak' (Confidence = 94,8%)

Struktur kaidah ini memberikan kejelasan audit yang transparan bagi analis kredit dalam menerbitkan surat rekomendasi persetujuan pembiayaan.

4. KESIMPULAN

Penelitian ini membuktikan bahwa integrasi algoritma C4.5 dan seleksi fitur Particle Swarm Optimization (PSO) sangat efektif dalam memprediksi kelayakan kredit calon debitur. Seleksi fitur PSO berhasil mereduksi 5 atribut redundan dan menyisakan 7 fitur prediktor optimal. Model C4.5 berbasis PSO mencatatkan performa terbaik dengan akurasi 89,17%, presisi 88,45%, recall 90,12%, F1-Score 89,28%, dan AUC-ROC 0,918, mengungguli Naive Bayes berbasis PSO (akurasi 84,67%).

Kaidah keputusan yang dihasilkan oleh model ini dapat langsung diadopsi sebagai Standard Operating Procedure (SOP) analitis otomatis pada sistem informasi pengajuan kredit lembaga keuangan mikro guna memangkas rasio kredit macet (NPL) secara berkelanjutan.

DAFTAR PUSTAKA

- Aktama, G. E. (2024). Penerapan data mining klasifikasi dalam peramalan kinerja dan risiko pada sistem perbankan daerah. *Jurnal Sistem Cerdas dan Informatika Terapan*, 9(2), 139–152. <https://doi.org/10.35870/jscit.v9i2.1604>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc.
- Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2), 103–130. <https://doi.org/10.1023/A:1007413511361>
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann. <https://doi.org/10.1016/C2018-0-02059-4>
- Handayani, R., & Hendrawan, B. (2020). Analisis faktor-faktor kelayakan kredit terhadap risiko kredit macet pada lembaga keuangan mikro. *Jurnal Riset Akuntansi dan Perbankan*, 14(1), 45–58.
- Kasmir. (2018). *Dasar-dasar perbankan* (Edisi revisi). PT RajaGrafindo Persada.
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, 4, 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers. <https://doi.org/10.1016/C2009-0-27660-9>
- Thomas, L. C., Crook, J. N., & Edelman, D. B. (2017). *Credit scoring and its applications* (2nd ed.). SIAM - Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611974560>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.
- Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2016). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, 20(4), 606–626. <https://doi.org/10.1109/TEVC.2015.2504420>