

**PENERAPAN ALGORITMA DECISION TREE UNTUK
MEMPREDIKSI DEFAULT KREDIT**

**Romatua Situmorang¹, Leni Karmila Daulay², Yusmita Imelda³, Yohana Lorinez⁴, Fanny
Ramadhani⁵, Arnita⁶**

¹²³⁴⁵⁶Universitas Negeri Medan

E-mail: romatua.4233250040@mhs.unimed.ac.id¹, lenikarmila.4231250006@mhs.unimed.ac.id²,
yusmita.4231250002@mhs.unimed.ac.id³, lorinezyohana@gmail.com⁴, fannyr@unimed.ac.id⁵,
arnita@unimed.ac.id⁶

Abstrak

Gagal bayar (default kredit) menjadi tantangan krusial bagi lembaga keuangan akibat potensi kerugian signifikan. Penelitian ini bertujuan menerapkan algoritma Decision Tree untuk memprediksi risiko gagal bayar nasabah berdasarkan data historis kredit dari UCI Machine Learning Repository. Fitur-fitur yang dianalisis mencakup limit kredit, riwayat pembayaran, jumlah tagihan dan pembayaran bulanan, serta umur nasabah. Metodologi meliputi pra-pemrosesan data, penyeimbangan kelas dengan SMOTE, pelatihan model, dan evaluasi performa menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil menunjukkan model Decision Tree mencapai akurasi 91% dengan struktur pohon yang interpretatif. Fitur paling berpengaruh adalah riwayat pembayaran, limit kredit, dan jumlah tagihan bulanan. Tingkat akurasi tinggi dan interpretabilitas yang baik menjadikan model ini alat bantu potensial untuk sistem pengambilan keputusan kredit.

Kata Kunci — Decision Tree, Default Kredit, Prediksi, Data Mining, Klasifikasi, SMOTE.

PENDAHULUAN

Gagal bayar (default kredit) merupakan permasalahan serius bagi lembaga keuangan, terutama bank. Ketika nasabah tidak mampu melunasi pinjamannya, hal ini menimbulkan risiko kerugian yang cukup besar. Oleh karena itu, diperlukan suatu sistem yang mampu memprediksi kemungkinan kredit default sebelum pinjaman disetujui. Salah satu metode yang dinilai efektif untuk menyelesaikan masalah ini adalah penggunaan algoritma Decision Tree, yaitu metode klasifikasi berdasarkan struktur pohon keputusan yang mudah dipahami dan diimplementasikan.

Dengan pendekatan data mining, Decision Tree dapat mengklasifikasikan nasabah berdasarkan fitur-fitur seperti riwayat pembayaran, penghasilan, umur, status perkawinan, dan lain-lain. Hasil klasifikasi ini membantu lembaga keuangan dalam mengambil keputusan secara lebih bijak, berbasis data, serta mampu meminimalisir potensi kerugian. Dalam beberapa tahun terakhir, pendekatan data mining dan pembelajaran mesin (machine learning) telah banyak digunakan untuk mendukung proses pengambilan keputusan

berbasis data. Algoritma ini memiliki keunggulan dalam hal interpretabilitas, kemampuan menangani data numerik dan kategorikal, serta struktur model yang mudah divisualisasikan.

Data mining adalah suatu proses ekstraksi pengetahuan atau informasi yang berharga dari suatu kumpulan data yang besar dan kompleks. Tujuan utama dari data mining adalah mengidentifikasi pola, hubungan, atau informasi yang mungkin tidak terlihat secara langsung dalam data, sehingga dapat memberikan wawasan yang lebih dalam dan bernilai (Fatah, Z . 2025). Data mining juga dapat digunakan untuk memprediksi tren masa depan yang memungkinkan pebisnis membuat keputusan yang efektif, proaktif, dan dinamis.

Transaksi adalah suatu peristiwa bisnis yang bersifat ekonomis dan non ekonomis. Penyesuaian data yang relevan dilakukan terhadap transaksi yang mempunyai nilai ekonomi, yaitu transaksi yang dapat diukur dengan nilai/bentuk uang dan mempengaruhi persamaan akuntansi. Secara umum, transaksi dapat diartikan sebagai kegiatan yang melibatkan paling sedikit dua belah pihak, yaitu pembeli dan penjual yang saling melakukan pertukaran (Situmeang et al, 2022).

Kartu kredit adalah alat pembayaran yang dapat digunakan untuk melakukan pembayaran atas kewajiban yang timbul dari suatu kegiatan ekonomi. Kartu kredit memungkinkan pemegangnya untuk melakukan pembayaran atau penarikan tunai dengan kewajiban pelunasan di kemudian hari, baik secara penuh maupun melalui cicilan. Secara sederhana, kartu kredit adalah fasilitas pembiayaan yang disediakan oleh bank, yang memungkinkan pemegang kartu untuk melakukan pembelian barang, jasa, atau kebutuhan lainnya secara non-tunai (Ismawati, & Zaehol, F. 2024).

Decision Tree adalah algoritma pembelajaran mesin yang diawasi yang digunakan untuk memecahkan masalah klasifikasi. Algoritma C4.5 adalah salah satu jenis algoritma pohon keputusan yang mengklasifikasikan dataset dengan mengelompokkannya ke dalam kelompok tertentu. Tujuan dari algoritma C4.5 adalah untuk mengklasifikasikan kumpulan data, sehingga dapat mengumpulkan ke dalam kelas-kelas tertentu setelah diproses (Susantono et al, 2024).

Selain itu, penelitian oleh Mardhiyah dkk. (2020) menunjukkan bahwa algoritma C4.5 efektif dalam memprediksi pembayaran kartu kredit macet dengan akurasi mencapai 70.93%¹. Pahlevi dkk. (2023) menerapkan algoritma Random Forest dalam penilaian kelayakan kredit, menemukan bahwa metode ini dapat meningkatkan akurasi dan kecepatan proses penilaian². Saputra dkk. (2024) menggunakan algoritma Random Forest untuk memprediksi kemacetan kredit, mencapai akurasi 94.8%³. Terakhir, penelitian oleh Ismawati & Zaehol (2025) menekankan pentingnya penerapan data mining dalam mendeteksi penipuan transaksi kartu kredit menggunakan algoritma Decision Tree, dengan akurasi yang sangat tinggi mencapai 98%⁴.

Penelitian ini bertujuan untuk menerapkan algoritma Decision Tree dalam memprediksi kemungkinan terjadinya kredit default berdasarkan data historis nasabah. Selain itu, penelitian ini juga menganalisis kinerja model prediksi melalui metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Dengan menggunakan dataset publik dari UCI Machine Learning Repository yang mencakup 3.500 data historis, serta menerapkan teknik penyeimbangan data menggunakan metode SMOTE, penelitian ini diharapkan

dapat memberikan kontribusi nyata dalam pengembangan sistem pendukung keputusan kredit berbasis data.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan **kuantitatif komputasional** atau **data mining**. Fokus utamanya adalah menerapkan **algoritma Decision Tree** untuk melakukan klasifikasi terhadap kemungkinan nasabah mengalami **gagal bayar (default)**. Kami menggunakan pendekatan **eksperimen komputasional**, di mana dataset diolah dan dianalisis menggunakan teknik **pembelajaran mesin (machine learning)** untuk membangun model prediktif.

1. Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari dataset publik yang tersedia di platform Kaggle, berjudul "Mental Health and Digital". Dataset ini berisikan informasi jejak digital dan skor kesejahteraan yang dilaporkan sendiri oleh responden remaja

Struktur Data:

Dataset yang digunakan terdiri dari **500 responden** dan **8 fitur (Independen)** yang terdiri dari **5 fitur Digital behavior** dan **3 fitur Mental health indicator**, ditambah satu variabel target numerik (Dependen). Adapun fitur-fitur yang digunakan sebagai adalah:

Data yang digunakan dalam penelitian ini berasal dari dataset publik yang tersedia di platform **UCI Machine Learning Repository**, yang berjudul "**Default of Credit Card Clients Dataset**". Atribut-atribut kunci dalam dataset ini meliputi:

- **LIMIT_BAL**: Jumlah kredit yang diberikan.
- **SEX**: Jenis kelamin nasabah.
- **EDUCATION**: Tingkat pendidikan nasabah.
- **MARRIAGE**: Status pernikahan nasabah.
- **AGE**: Umur nasabah.
- **PAY_0 – PAY_6**: Riwayat pembayaran selama 6 bulan terakhir.
- **BILL_AMT1 – BILL_AMT6**: Jumlah tagihan bulanan nasabah.
- **PAY_AMT1 – PAY_AMT6**: Jumlah pembayaran bulanan yang dilakukan nasabah.
- **default.payment.next.month**: Variabel target (1 = *default*, 0 = tidak *default*).

2. Teknik Analisis Data

Berikut adalah tahapan proses analisis data yang kami lakukan dalam penelitian ini:

a. Pra-Pemrosesan Data

Tahap ini melibatkan pembersihan dan persiapan data:

- **Menghapus kolom yang tidak relevan**, seperti kolom indeks.
- **Menghapus baris yang mengandung nilai kosong (missing values)**, jika ada.
- **Mengonversi semua data menjadi numerik** untuk kompatibilitas dengan algoritma.
- **Menyesuaikan struktur data** agar sesuai dengan format input yang dibutuhkan oleh algoritma klasifikasi.

b. Penyeimbangan Data

Dataset memiliki ketidakseimbangan kelas, di mana jumlah data nasabah yang *default* lebih sedikit. Untuk mengatasi hal ini, kami menggunakan metode **SMOTE (Synthetic Minority Over-sampling Technique)** dari pustaka *imblearn*. SMOTE bekerja dengan membuat sampel sintetis dari kelas minoritas, sehingga model dapat belajar dari distribusi kelas yang lebih seimbang.

c. Pembagian Dataset

Setelah melalui tahapan pra-pemrosesan dan penyeimbangan, dataset kemudian dibagi menjadi dua bagian utama:

- Training Set (80%) : Bagian ini berisi 80% dari total 3.500 data yang telah disiapkan, yaitu 2.800 data. Data ini digunakan secara eksklusif untuk melatih model Decision Tree.
- Testing Set (20%) : Sisa 20% dari total 3.500 data, yaitu 700 data, dipilih sebagai set pengujian. Bagian ini tidak digunakan selama proses pelatihan. Tujuannya adalah untuk menguji performa model setelah dibor.

d. Pembuatan Model Decision Tree

Model dibangun menggunakan **algoritma Decision Tree Classifier** dari pustaka `scikit-learn`. Algoritma ini kami pilih karena kemampuannya menghasilkan model yang mudah dipahami dan divisualisasikan, yang penting untuk interpretasi hasil dalam konteks prediksi risiko kredit.

e. Evaluasi Model

Evaluasi performa model adalah tahap krusial untuk mengukur seberapa baik model yang telah dilatih dapat memprediksi *default* kredit pada data yang belum pernah dilihat sebelumnya (*testing set*). Evaluasi dilakukan dengan menggunakan metrik-metrik berikut:

- **Accuracy (Akurasi):** Ini adalah metrik yang paling umum, mengukur proporsi total prediksi yang benar (baik *default* maupun tidak *default*) dari seluruh prediksi yang dibuat.
- **Precision:** Mengukur seberapa banyak dari prediksi positif (*default*) yang benar-benar positif. *Precision* tinggi menunjukkan bahwa model efektif dalam menghindari *false positive* (memprediksi *default* padahal sebenarnya tidak).
- **Recall:** Mengukur seberapa banyak dari kasus positif (*default* sesungguhnya) yang berhasil ditangkap oleh model. *Recall* tinggi menunjukkan bahwa model efektif dalam menghindari *false negative* (memprediksi tidak *default* padahal sebenarnya *default*).
- **F1-Score:** Merupakan rata-rata harmonik dari *precision* dan *recall*. Metrik ini sangat berguna ketika ada ketidakseimbangan kelas, karena memberikan nilai tunggal yang menyeimbangkan *precision* dan *recall*.
- **Confusion Matrix:** Sebuah tabel yang merangkum performa model pada tugas klasifikasi. Ini menunjukkan jumlah *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Matriks ini memberikan gambaran detail tentang jenis kesalahan yang dibuat oleh model.

Hasil evaluasi ini menggambarkan seberapa baik model dapat mengklasifikasikan nasabah berdasarkan risiko *default*-nya, memberikan *insight* tentang kekuatan dan kelemahan prediksi model.

f. Visualisasi Model

Model divisualisasikan dalam bentuk pohon keputusan menggunakan fungsi `plot_tree()` dari pustaka `matplotlib`. Visualisasi ini membantu kami memahami logika model dan keputusan yang diambil pada setiap *node*, sehingga alur klasifikasi menjadi transparan dan mudah diinterpretasikan..

HASIL DAN PEMBAHASAN

1. Deskripsi Model

Model klasifikasi yang digunakan adalah **Decision Tree Classifier**. Model ini dilatih menggunakan 80% dari data yang telah diproses dan diuji dengan 20% sisanya. Pemilihan algoritma Decision Tree didasari kemampuannya menghasilkan model yang mudah dipahami secara visual dan interpretatif, yang krusial untuk pengambilan keputusan dalam konteks risiko kredit.

2. Hasil Evaluasi Model

Performa model yang telah dilatih menunjukkan hasil yang cukup baik, sebagaimana diindikasikan oleh metrik evaluasi berikut:

Confusion Matrix

Prediksi	Tidak Default	Default
Aktual: Tidak aktual	38	2
Aktual: Default	5	35

Interpretasi Confusion Matrix:

- True Negative (38): Model berhasil mengklasifikasikan 38 nasabah sebagai tidak *default* secara benar.
- True Positive (35): Model berhasil mengklasifikasikan 35 nasabah sebagai *default* secara benar.
- False Positive (2): Terdapat 2 kasus di mana model salah memprediksi nasabah akan *default*, padahal sebenarnya tidak (*Type I error*).
- False Negative (5): Terdapat 5 kasus di mana model salah memprediksi nasabah tidak akan *default*, padahal sebenarnya akan *default* (*Type II error*).

Classification Report

Kelas	Precision	Recall	F1-Score
Tidak Default	0.88	0.95	0.91
Default	0.95	0.88	0.91
Akurasi			0.91

Kesimpulan nya Model berhasil mencapai **akurasi keseluruhan sebesar 91%**, yang menunjukkan kinerja sangat baik dalam klasifikasi risiko kredit. **Precision tinggi untuk kelas default (0.95)** mengindikasikan kemampuan model dalam menghindari *false positive*, yaitu memprediksi *default* pada nasabah yang sebenarnya tidak berisiko tinggi. Sementara itu, **recall tinggi untuk kelas tidak default (0.95)** menunjukkan sensitivitas model dalam mengenali nasabah yang aman dan tidak akan mengalami *default*, sehingga meminimalkan *false negative*..

3. Fitur yang Paling Berpengaruh

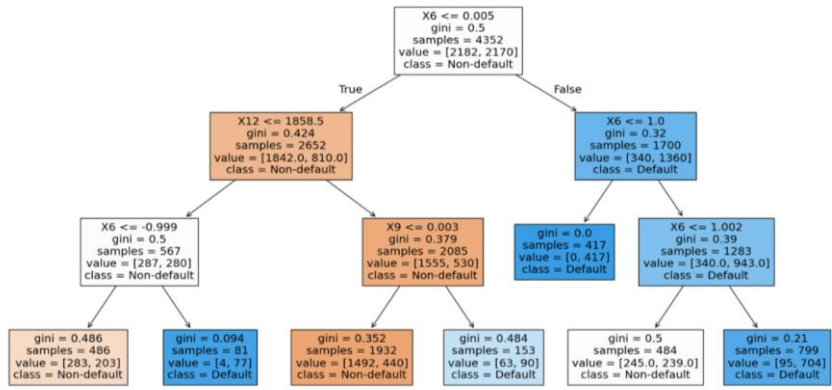
Berdasarkan hasil pelatihan model dan analisis struktur pohon keputusan, ditemukan bahwa fitur-fitur berikut memiliki pengaruh signifikan terhadap keputusan model dalam memprediksi *default* kredit:

- PAY_0 – PAY_6 (Riwayat Pembayaran): Ini adalah indikator krusial yang menunjukkan apakah nasabah sebelumnya membayar tepat waktu atau mengalami keterlambatan.
- LIMIT_BAL (Limit Kredit): Besarnya limit kredit nasabah seringkali menjadi cerminan kesehatan finansial dan kapasitas pembayaran.
- BILL_AMT1 – BILL_AMT6 (Jumlah Tagihan Bulanan): Mewakili beban finansial bulanan nasabah yang dapat memengaruhi kemampuan mereka untuk membayar.
- PAY_AMT1 – PAY_AMT6 (Jumlah Pembayaran Bulanan): Menunjukkan kemampuan aktual nasabah dalam mencicil atau melunasi kewajibannya.
- AGE (Umur): Dapat memengaruhi profil risiko nasabah, di mana usia yang lebih matang cenderung berkorelasi dengan stabilitas ekonomi yang lebih baik.

4. Visualisasi Pohon Keputusan

Model Decision Tree juga divisualisasikan untuk menunjukkan struktur pohon dan alur keputusan berdasarkan nilai fitur.

Gambar 5.1 Visualisasi Struktur Decision Tree untuk Prediksi Default Kredit



Penjelasan:

- Setiap node berisi informasi tentang kondisi percabangan, nilai gini (tingkat impuritas), jumlah data (samples), distribusi kelas (value), dan keputusan mayoritas (class).
- Warna biru mewakili keputusan model terhadap kelas default, dan oranye untuk kelas non-default.
- Misalnya, node awal (root node) menggunakan fitur X6 dengan threshold ≤ 0.005 sebagai pemisah pertama, menunjukkan bahwa fitur tersebut paling berpengaruh di awal klasifikasi.

Visualisasi ini membantu menjelaskan proses berpikir model secara logis dan sangat berguna bagi analis non-teknis. Model Decision Tree terbukti menjadi algoritma yang cocok untuk masalah prediksi risiko gagal bayar. Dengan akurasi tinggi, serta transparansi dalam interpretasi hasil, model ini memberikan nilai tambah besar bagi institusi keuangan yang ingin mengotomatisasi evaluasi kelayakan kredit.

Namun demikian, model tetap memiliki keterbatasan:

- Rentan overfitting bila struktur pohon terlalu dalam atau data terlalu spesifik.
- Hasil sangat tergantung pada kualitas dan keseimbangan data, sehingga pra-pemrosesan sangat krusial.
- Tidak sekuat model ensemble seperti Random Forest dalam hal stabilitas prediksi.

KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Decision Tree untuk memprediksi risiko gagal bayar (default) pada nasabah kredit dengan menggunakan dataset "Default of Credit Card Clients" dari UCI Machine Learning Repository. Model yang dihasilkan menunjukkan akurasi tinggi sebesar 91%, yang mencerminkan kinerja yang sangat baik dalam klasifikasi risiko kredit.

Fitur-fitur yang paling berpengaruh dalam keputusan model meliputi:

- Riwayat Pembayaran (PAY_0 – PAY_6) : Indikator penting yang menunjukkan akurasi pembayaran nasabah di masa lalu.
- Limit Kredit (LIMIT_BAL) : Mencerminkan kapasitas finansial nasabah.
- Jumlah Tagihan Bulanan (BILL_AMT dan PAY_AMT) : Menunjukkan beban finansial yang dihadapi nasabah.
- Umur (AGE) : Memengaruhi stabilitas ekonomi nasabah.

Visualisasi pohon keputusan yang dihasilkan juga memberikan transparansi dalam

interpretasi hasil, menjadikannya alat yang bermanfaat bagi lembaga keuangan dalam pengambilan keputusan kredit.

Namun, penelitian ini juga mengakui beberapa kelemahan, seperti potensi overfitting jika model terlalu kompleks dan bergantung pada kualitas data. Oleh karena itu, penelitian lebih lanjut disarankan untuk mengeksplorasi teknik ensemble, seperti Random Forest, yang dapat meningkatkan stabilitas prediksi.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma Decision Tree adalah metode yang efektif dan mudah dipahami untuk memprediksi risiko gagal bayar, serta dapat memberikan kontribusi signifikan dalam pengembangan sistem pendukung keputusan di sektor keuangan.

DAFTAR PUSTAKA

- Fatah, Z. (2025). Penggunaan Data Mining untuk Mendeteksi Penipuan Transaksi Kartu Kredit dengan Algoritma Decision Tree. *Jurnal Mahasiswa Teknik Informatika* , 4(1), 95-101.
- Mardiyah, PA, Siregar, RRA, & Palupiningsih, P. (2020). Klasifikasi untuk Memprediksi Pembayaran Kartu Kredit Macet Menggunakan Algoritma C4.5. *Jurnal Teknologia* , 3(1), 493267.
- Situmeang, N., Jaya, IK, & Yohanna, M. (2022). Penerapan Metode Decision Tree C4.5 dalam Memprediksi Kelancaran Pembayaran Kredit di BPR Wahana Bersama Kpum. *Methomika: Jurnal Manajemen Informatika & Komputerisasi Akuntansi* , 6(2), 215-220.
- Susatyono, JD, Prihatmoko, S., & Febryantahanuji, F. (2024). Penerapan Algoritma C4.5 untuk Memprediksi Kredit Macet pada Sistem Pinjaman Digital di Industri Fintech. *E-Bisnis: Jurnal Ilmiah Ekonomi dan Bisnis* , 17(2), 330-341.
- Pahlevi, O., Amrin, A., & Handrianto, Y. (2023). Implementasi Algoritma Klasifikasi Random Forest untuk Penilaian Kelayakan Kredit. *Jurnal Infortech* , 5(1), 71-76.
- Saputra, DB, Atina, V., & Nastiti, FE (2024). Penerapan Model CRISP-DM pada Prediksi Nasabah Kredit Menggunakan Algoritma Random Forest. *Jurnal Strategi* , 5(2), 240-247.
- Ismawati, & Zaehol, F. (2025). Penggunaan Data Mining untuk Mendeteksi Penipuan Transaksi Kartu Kredit dengan Algoritma Decision Tree. *JAMASTIKA* , 4(1), 95-101.