

**KLASIFIKASI KELAYAKAN KREDIT MENGGUNAKAN
GRADIENT BOOSTING DAN L1-BASED FEATURE
SELECTION**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ABDURRAHMAN RIDHO

21.11.4341

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**KLASIFIKASI KELAYAKAN KREDIT MENGGUNAKAN
GRADIENT BOOSTING DAN L1-BASED FEATURE
SELECTION**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ABDURRAHMAN RIDHO

21.11.4341

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI KELAYAKAN KREDIT MENGGUNAKAN GRADIENT
BOOSTING DAN L1-BASED FEATURE SELECTION**

yang disusun dan diajukan oleh

Abdurrahman Ridho

21.11.4341

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Agustus 2025

Dosen Pembimbing,



I Made Artha Agastya, S.T., M.Eng., Ph.D.

NIK. 190302352

HALAMAN PENGESAHAN

SKRIPSI

**KLASIFIKASI KELAYAKAN KREDIT MENGGUNAKAN GRADIENT
BOOSTING DAN L1-BASED FEATURE SELECTION**

yang disusun dan diajukan oleh

Abdurrahman Ridho

21.11.4341

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Agustus 2025

Susunan Dewan Penguji

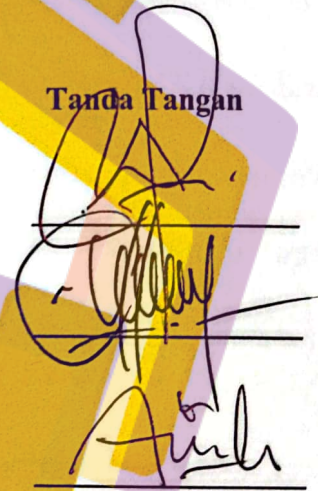
Nama Penguji

Tanda Tangan

Dr. Andi Sunyoto, S.Kom., M.Kom.
NIK. 190302052

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

I Made Artha Agastya, S.T., M.Eng., Ph.D.
NIK. 190302352



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Agustus 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

...

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Abdurrahman Ridho

NIM : 21.11.4341

Menyatakan bahwa Skripsi dengan judul berikut:

KLASIFIKASI KELAYAKAN KREDIT MENGGUNAKAN GRADIENT BOOSTING DAN L1-BASED FEATURE SELECTION

Dosen Pembimbing : I Made Artha Agastya, S.T., M.Eng., Ph.D.,

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat **orang** lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 21 Agustus 2025

Yang Menyatakan,



Abdurrahman Ridho

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tesis yang berjudul “KLASIFIKASI KELAYAKAN KREDIT MENGGUNAKAN GRADIENT BOOSTING DAN L1-BASED FEATURE SELECTION” dengan baik. Tesis ini disusun sebagai salah satu syarat untuk menyelesaikan program studi Informatika di Universitas Amikom Yogyakarta.

Proses penyusunan tesis ini tidak akan berjalan lancar tanpa adanya bimbingan, dukungan, dan doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang tulus kepada:

1. Bapak I Made Artha Agastya, S.T., M.Eng., Ph.D., selaku Dosen Pembimbing. Penulis sangat berterima kasih atas segala bimbingan, arahan, waktu, serta semangat yang telah diberikan selama proses penyusunan tesis ini. Dukungan dan masukan dari Bapak sangat berarti bagi penulis dalam mengatasi berbagai tantangan dan menyelesaikan penelitian ini.
2. Keluarga tercinta, terutama Ayah dan Ibu yang senantiasa memberikan dukungan moril, materiel, serta doa yang tiada henti.
3. Sahabat-sahabat dan rekan-rekan seperjuangan yang telah memberikan semangat dan menjadi teman diskusi yang baik.

Penulis menyadari bahwa tesis ini masih jauh dari kata sempurna. Oleh karena itu, segala bentuk kritik dan saran yang membangun akan penulis terima dengan lapang dada demi perbaikan di masa mendatang. Semoga tesis ini dapat memberikan manfaat dan kontribusi bagi perkembangan ilmu pengetahuan.

Yogyakarta, 17 Agustus 2025

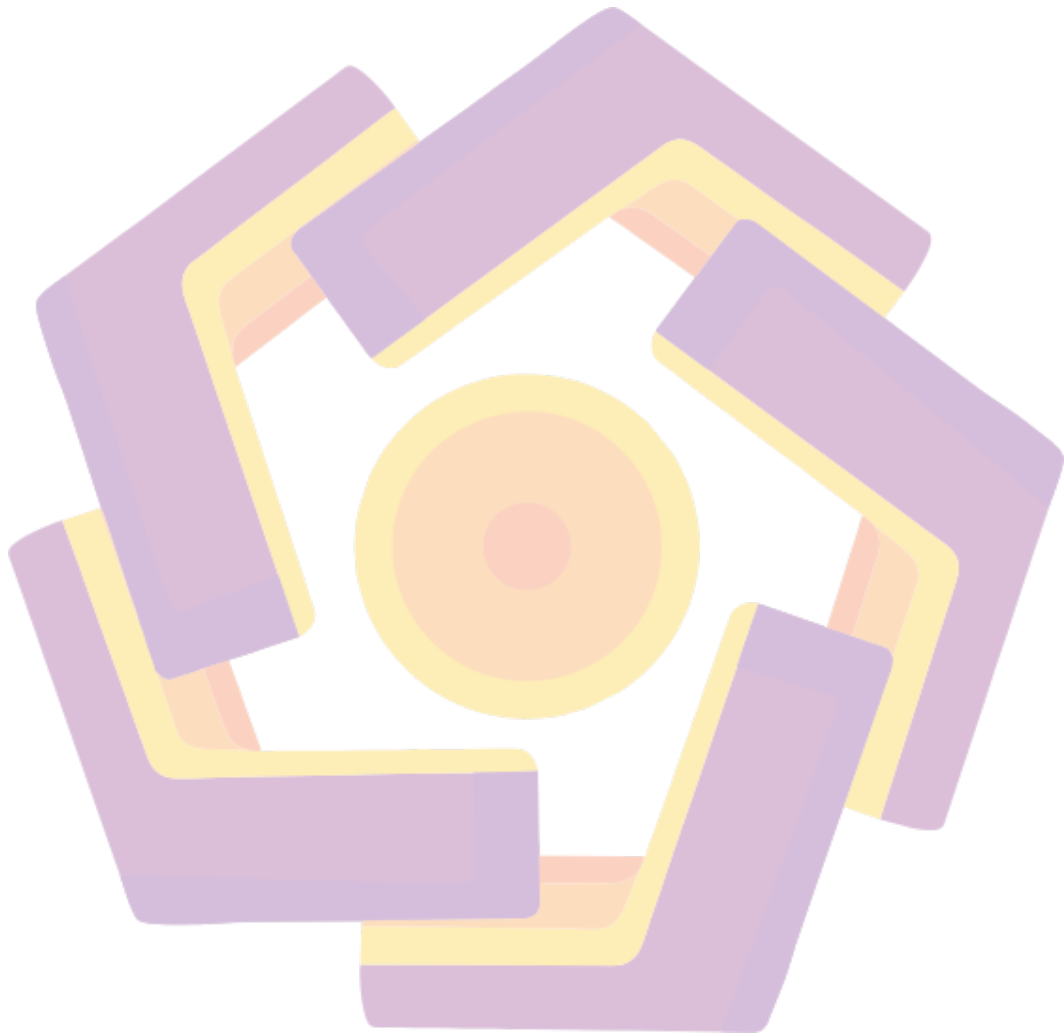
Abdrrahman Ridho

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMBANG DAN SINGKATAN	xi
DAFTAR ISTILAH	xii
INTISARI	xiii
<i>ABSTRACT</i>	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	14
2.2.1 Pinjaman	14
2.2.2 Risiko kredit	14
2.2.3 Kelayakan kredit	14
2.2.4 Klasifikasi	14
2.2.5 Machine Learning	15
2.2.6 SMOTE	15

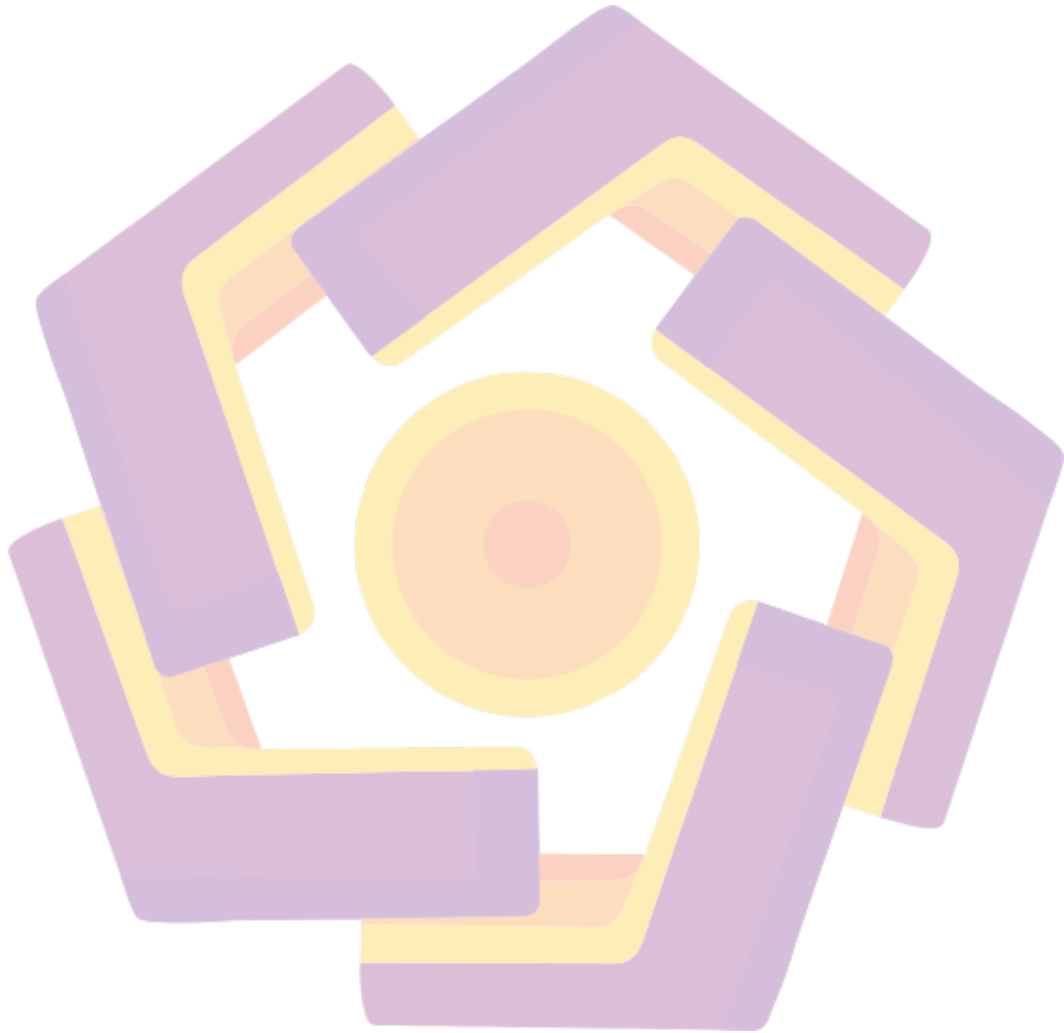
2.2.7	Grid Search	15
2.2.8	Feature Selection	15
2.2.9	K-Fold Cross Validation	16
2.2.10	Gradient Boosting	16
BAB III METODE PENELITIAN		18
3.1	Objek Penelitian	18
3.2	Alur Penelitian	18
3.2.1	Pengumpulan Data	20
3.2.2	Preprocessing Data	20
3.2.3	Feature Selection	21
3.2.4	Splitting Data	21
3.2.5	Implementasi Model Gradient Boost	21
3.2.6	Analisis	22
3.3	Alat dan Bahan	22
3.3.1	Alat	22
3.3.2	Bahan	23
BAB IV HASIL DAN PEMBAHASAN		24
4.1	Pengumpulan Data	24
4.2	Exploratory Data Analysis (EDA)	25
4.2.1	Heatmap	25
4.2.2	Analisis Distribusi Kelas Target	26
4.2.3	Analisis pada Fitur Kategorikal	27
4.3	Preprocessing Data	29
4.3.1	Cleaning	29
4.3.2	Encoding	30
4.3.3	Smote	32
4.4	Feature Selection	34
4.5	Splitting Data	38
4.6	Analisis	39
4.6.1	Evaluasi K-Fold	39
4.6.2	Confusion Matrix	41
4.6.3	Hasil Evaluasi	41

BAB V PENUTUP	45
5.1 Kesimpulan	45
5.2 Saran	45
REFERENSI	47



DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	8
Tabel 4. 1 Hasil evaluasi	41

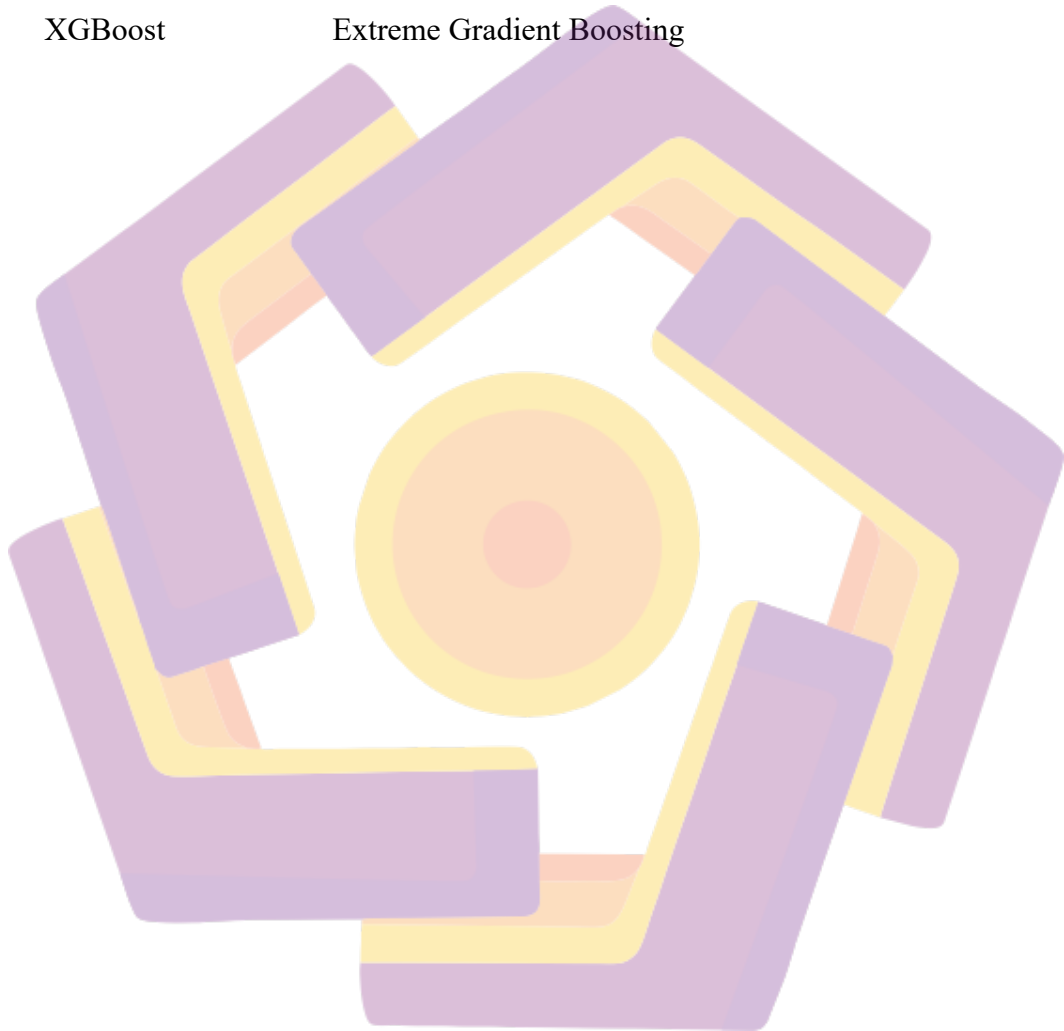


DAFTAR GAMBAR

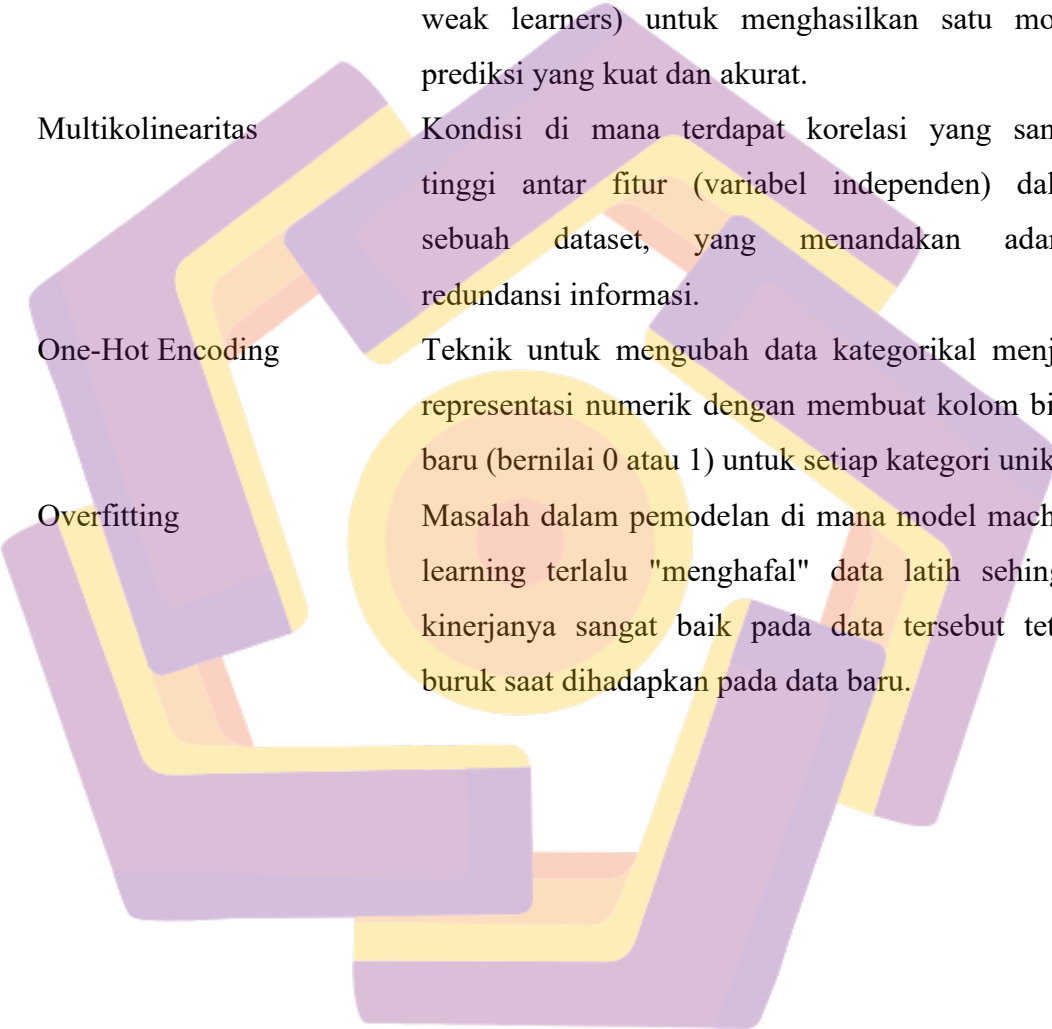
Gambar 3. 1 Alur Penelitian	19
Gambar 4. 1 Heatmap	25
Gambar 4. 2 Distribusi Kelas pada Variabel Target Loan Approved	26
Gambar 4. 3 Visualisasi Fitur Kategorikal	27
Gambar 4. 4 ApplicationDate	29
Gambar 4. 5 Hasil Setelah Encoding	31
Gambar 4. 6 Sebelum Smote	33
Gambar 4. 7 Sesudah Smote	33
Gambar 4. 8 Feature Yang Digunakan Model tanpa seleksi fitur.....	34
Gambar 4. 9 Feature Yang Digunakan Korelasi 0,9	35
Gambar 4. 10 Feature Yang Digunakan Korelasi 0,7	36
Gambar 4. 11 Feature Yang Digunakan Korelasi 0,5	37
Gambar 4. 12 Feature Yang Digunakan RFE	37
Gambar 4. 13 Feature Yang Digunakan SelectKBest.....	38
Gambar 4. 14 Feature Yang Digunakan L1-Based	38
Gambar 4. 15 Evaluasi K-Fold	39
Gambar 4. 16 Confusion Matrix	41

DAFTAR LAMBANG DAN SINGKATAN

RFE	Recursive Feature Elimination
SMOTE	Synthetic Minority Over-sampling Technique
XGBoost	Extreme Gradient Boosting



DAFTAR ISTILAH



Ensemble Learning	Metode dalam machine learning yang menggabungkan beberapa model sederhana (disebut weak learners) untuk menghasilkan satu model prediksi yang kuat dan akurat.
Multikolinearitas	Kondisi di mana terdapat korelasi yang sangat tinggi antar fitur (variabel independen) dalam sebuah dataset, yang menandakan adanya redundansi informasi.
One-Hot Encoding	Teknik untuk mengubah data kategorikal menjadi representasi numerik dengan membuat kolom biner baru (bernilai 0 atau 1) untuk setiap kategori unik.
Overfitting	Masalah dalam pemodelan di mana model machine learning terlalu "menghafal" data latih sehingga kinerjanya sangat baik pada data tersebut tetapi buruk saat dihadapkan pada data baru.

INTISARI

Penyaluran kredit yang tepat sasaran merupakan faktor penting dalam mengurangi risiko kredit bermasalah dan meningkatkan efisiensi lembaga keuangan. Namun, proses penilaian kelayakan kredit sering kali menghadapi tantangan akibat kompleksitas data peminjam. Oleh karena itu, diperlukan metode klasifikasi yang mampu memberikan hasil prediksi dengan tingkat akurasi yang tinggi.

Penelitian ini bertujuan untuk melakukan klasifikasi persyaratan kelayakan peminjaman kredit menggunakan algoritma Gradient Boosting dengan penerapan beberapa metode seleksi fitur. Empat metode yang digunakan yaitu Korelasi, SelectKBest, L1-Based (Lasso), dan Recursive Feature Elimination (RFE). Dataset yang digunakan merupakan data publik dari platform Kaggle.

Hasil penelitian menunjukkan bahwa metode seleksi fitur memberikan pengaruh signifikan terhadap kinerja model. Metode L1-Based menghasilkan performa terbaik dengan akurasi sebesar 97,07% dan F1-Score sebesar 97,08%, melampaui baseline tanpa seleksi fitur yang memiliki akurasi 95,40%. Sementara itu, metode korelasi 0,9 dan RFE juga memberikan hasil yang mendekati baseline, sedangkan SelectKBest menunjukkan performa terendah. Dengan demikian, metode L1-Based terbukti paling efektif dalam meningkatkan akurasi klasifikasi kelayakan peminjaman kredit.

Kata kunci: Gradient Boosting, Seleksi Fitur, Klasifikasi Kredit, Machine Learning, Risiko Kredit

ABSTRACT

Accurate credit approval assessment plays a crucial role in reducing the risk of non-performing loans and improving the efficiency of financial institutions. However, the process of evaluating borrower eligibility often faces challenges due to the complexity of financial data. Therefore, a robust classification method is required to produce high-accuracy predictions.

This study aims to classify credit eligibility requirements using the Gradient Boosting algorithm combined with several feature selection methods. Four methods were applied: Correlation, SelectKBest, L1-Based (Lasso), and Recursive Feature Elimination (RFE). The dataset was obtained from a publicly available source on the Kaggle platform.

The results show that feature selection significantly affects model performance. The L1-Based method achieved the best performance with an accuracy of 97.07% and an F1-Score of 97.08%, surpassing the baseline model without feature selection, which achieved 95.40% accuracy. Meanwhile, Correlation (0.9) and RFE methods also provided competitive results, whereas SelectKBest produced the lowest performance. Therefore, the L1-Based method is proven to be the most effective in improving credit eligibility classification accuracy.

Keyword: Gradient Boosting, Feature Selection, Credit Classification, Machine Learning, Credit Risk