

**PERBANDINGAN ALGORITMA XGBOOST DAN C4.5
DALAM MEMPREDIKSI SERTA MENGLASIFIKASIKAN
RISIKO KREDIT BERDASARKAN DATA PEMOHON UNTUK
KELAYAKAN PINJAMAN**

SKRIPSI

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



disusun oleh

RASTRA ARDIANSYAH PORA

21.11.4077

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PERBANDINGAN ALGORITMA XGBOOST DAN C4.5
DALAM MEMPREDIKSI SERTA MENGLASIFIKASIKAN
RISIKO KREDIT BERDASARKAN DATA PEMOHON UNTUK
KELAYAKAN PINJAMAN**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

RASTRA ARDIANSYAH PORA

21.11.4077

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN ALGORITMA XGBOOST DAN C4.5 DALAM
MEMPREDIKSI SERTA MENGLASIFIKASIKAN RISIKO KREDIT
BERDASARKAN DATA PEMOHON UNTUK KELAYAKAN PINJAMAN**

yang disusun dan diajukan oleh

Rastra Ardiansyah Pora

21.11.4077

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Januari 2025

Dosen Pembimbing,



Dina Maulina, M.Kom

NIK. 190302250

HALAMAN PENGESAHAN

SKRIPSI

PERBANDINGAN ALGORITMA XGBOOST DAN C4.5 DALAM
MEMPREDIKSI SERTA MENGLASIFIKASIKAN RISIKO KREDIT
BERDASARKAN DATA PEMOHON UNTUK KELAYAKAN PINJAMAN

yang disusun dan diajukan oleh

Rastra Ardiansyah Pora

21.11.4077

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Januari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Wahid Miftahul Ashari, S.Kom., M.T
NIK. 190302452

Alfie Nur Rahmi, S.Kom., M.Kom
NIK. 190302240

Dina Maulina, S.Kom., M.Kom
NIK. 190302250

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Januari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Rastra Ardiansyah Pora
NIM : 21.11.4077

Menyatakan bahwa Skripsi dengan judul berikut:

Perbandingan Algoritma XGBoost dan C4.5 dalam Memprediksi serta Mengklasifikasikan Risiko Kredit Berdasarkan Data Pemohon untuk Kelayakan Pinjaman

Dosen Pembimbing : Dina Maulina, M.Kom

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, 22 Januari 2025

Yang Menyatakan,



Rastra Ardiansyah Pora

HALAMAN PERSEMBAHAN

Segala puji syukur panjatkan kehadiran Allah SWT atas limpahan rahmat, hidayah, dan kekuatan-Nya sehingga penulis dapat menyelesaikan skripsi ini sebagai salah satu syarat untuk mencapai gelar Sarjana pada Program Studi Informatika.

Dengan penuh rasa terima kasih dan hormat, skripsi ini penulis persembahkan kepada:

1. Kedua orang tua tercinta dan keluarga penulis yang selalu memberikan doa, dukungan, kasih sayang, serta semangat tanpa henti.
2. Dosen pembimbing Ibu Dina Maulina, yang dengan sabar memberikan bimbingan, arahan, serta ilmu yang berharga dalam penyusunan skripsi ini.
3. Dosen penguji yang telah memberikan kritik dan saran yang membangun demi kesempurnaan penelitian ini.
4. Terima kasih kepada seluruh dosen yang telah berbagi ilmu dan pengalaman selama masa perkuliahan.
5. Teman-teman seperjuangan yang selalu memberikan dukungan, kebersamaan, dan semangat dalam menyelesaikan perjalanan akademik ini.

Semoga segala kebaikan yang telah diberikan mendapat balasan yang setimpal.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika.

Dalam proses penyusunan skripsi ini, penulis menerima banyak dukungan, bimbingan, serta motivasi dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan terima kasih, penulis menyampaikan penghargaan yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Dina Maulina, S.Kom., M.Kom selaku Dosen Pembimbing yang dengan penuh kesabaran telah memberikan arahan, bimbingan, serta masukan yang sangat berarti dalam penyusunan skripsi ini.
4. Wahid Miftahul Ashari, S.Kom., M.T dan Ale Nur Rahmi, S.Kom., M.Kom selaku Tim Dosen Penguji yang telah memberikan kritik, saran, serta evaluasi yang sangat berharga dalam penyempurnaan skripsi ini.
5. Seluruh dosen Universitas Amikom Yogyakarta, yang telah memberikan ilmu dan wawasan yang sangat bermanfaat selama masa perkuliahan.
6. Kedua orang tua tercinta dan keluarga yang selalu mendoakan, memberikan dukungan moral, material, serta kasih sayang yang tak terbatas.
7. Teman-teman seperjuangan yang telah memberikan semangat, kebersamaan, dan dukungan selama proses penyusunan skripsi ini.

Yogyakarta, 22 Januari 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
INTISARI.....	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Studi Literatur.....	6
2.2 Dasar Teori.....	17
2.2.1 Kreditur dan Resiko Kredit.....	17
2.2.2 Data Mining.....	18
2.2.3 <i>Knowledge Discovery in Data (KDD)</i>	19
2.2.4 Pohon Keputusan (<i>Decision Tree</i>).....	21
2.2.5 C4.5.....	23
2.2.6 Metode <i>Ensemble Machine Learning</i>	25

2.2.7	XGBoost.....	26
2.2.8	<i>Confusion Matrix</i>	28
2.2.9	<i>GridSearchCV</i>	29
2.2.10	<i>K-Fold Cross Validation</i>	30
BAB III METODE PENELITIAN.....		32
3.1	Objek Penelitian.....	32
3.2	Alur Penelitian.....	34
3.2.1	Pengambilan Data.....	35
3.2.2	<i>Data Preprocessing</i>	36
3.2.3	<i>Data Transformation</i>	37
3.2.4	Pemodelan dan Evaluasi.....	40
3.3	Alat dan Bahan.....	50
3.3.1	Data Penelitian.....	50
3.3.2	Alat/instrumen.....	52
BAB IV HASIL DAN PEMBAHASAN.....		53
4.1	Hasil Data <i>Preprocessing</i>	53
4.2	Hasil Data <i>Transformation</i>	55
4.3	Hasil Pemodelan dan Evaluasi Algoritma XGBoost.....	58
4.4	Hasil Pemodelan Algoritma XGBoost GridSearchCV.....	66
4.5	Hasil Pemodelan Algoritma C4.5.....	75
4.6	Hasil Pemodelan Algoritma C4.5 GridSearchCV.....	83
4.7	Perbandingan Hasil.....	91
4.8	Ringkasan <i>Prototype</i> Model XGBoost dan C4.5.....	95
BAB V PENUTUP.....		98
5.1	Kesimpulan.....	98
5.2	Saran.....	99
REFERENSI.....		101
LAMPIRAN.....		102

DAFTAR TABEL

Tabel 2.1	Keaslian Penelitian.....	11
Tabel 3.1	Keterangan Atribut.....	32
Tabel 3.2	Spesifikasi Perangkat Keras.....	52
Tabel 3.3	Spesifikasi Perangkat Lunak.....	52
Tabel 4.1	Hasil Data <i>Validation</i> (Cek nilai null dan duplikat).....	53
Tabel 4.2	Hasil Data <i>Cleaning</i> (Perubahan nama atribut).....	54
Tabel 4.3	Hasil Data <i>Cleaning</i> (Penghapusan kolom atribut).....	54
Tabel 4.4	Hasil atribut setelah Data <i>Preprocessing</i>	55
Tabel 4.5	Hasil <i>Scaling</i> (Perubahan nilai atribut).....	55
Tabel 4.6	Hasil Label <i>Encoding</i> (Perubahan kategori atribut).....	56
Tabel 4.7	Hasil <i>Oversampling</i> (Menyeimbangkan data).....	56
Tabel 4.8	Hasil atribut setelah Data <i>Preprocessing</i> dan <i>Transformation</i>	57
Tabel 4.9	Hasil Rasio <i>Split Data</i>	58
Tabel 4.10	Hasil data latih dan uji (XGBoost).....	59
Tabel 4.11	K 10 <i>Fold Cross Validation</i> 90:10 (XGBoost).....	60
Tabel 4.12	K 10 <i>Fold Cross Validation</i> 80:20 (XGBoost).....	60
Tabel 4.13	K 10 <i>Fold Cross Validation</i> 70:30 (XGBoost).....	61
Tabel 4.14	K 10 <i>Fold Cross Validation</i> 60:40 (XGBoost).....	61
Tabel 4.15	Performa data uji (XGBoost).....	62
Tabel 4.16	Nilai ROC (AUC) data uji (XGBoost).....	63
Tabel 4.17	<i>Classification Report</i> 90:10 data uji (XGBoost).....	65
Tabel 4.18	<i>Classification Report</i> 80:20 data uji (XGBoost).....	65
Tabel 4.19	<i>Classification Report</i> 70:30 data uji (XGBoost).....	66
Tabel 4.20	<i>Classification Report</i> 60:40 data uji (XGBoost).....	66
Tabel 4.21	Jangkauan parameter GridSearchCV (XGBoost).....	66
Tabel 4.22	Hasil parameter terbaik GridSearchCV (XGBoost).....	67
Tabel 4.23	Hasil GridSearchCV data latih dan uji (XGBoost).....	68
Tabel 4.24	GridSearchCV K 10 <i>Fold Cross Validation</i> 90:10 (XGBoost).....	69
Tabel 4.25	GridSearchCV K 10 <i>Fold Cross Validation</i> 80:20 (XGBoost).....	69
Tabel 4.26	GridSearchCV K 10 <i>Fold Cross Validation</i> 70:30 (XGBoost).....	70
Tabel 4.27	GridSearchCV K 10 <i>Fold Cross Validation</i> 60:40 (XGBoost).....	70
Tabel 4.28	Performa GridSearchCV data uji (XGBoost).....	71
Tabel 4.29	Nilai ROC (AUC) GridSearchCV data uji (XGBoost).....	72
Tabel 4.30	<i>Classification Report</i> GridSearchCV 90:10 data uji (XGBoost).....	74
Tabel 4.31	<i>Classification Report</i> GridSearchCV 80:20 data uji (XGBoost).....	74
Tabel 4.32	<i>Classification Report</i> GridSearchCV 70:30 data uji (XGBoost).....	75
Tabel 4.33	<i>Classification Report</i> GridSearchCV 60:40 data uji (XGBoost).....	75
Tabel 4.34	Hasil Rasio <i>Split Data</i>	76
Tabel 4.35	Hasil data latih dan uji (C4.5).....	76
Tabel 4.36	K 10 <i>Fold Cross Validation</i> 90:10 (C4.5).....	77
Tabel 4.37	K 10 <i>Fold Cross Validation</i> 80:20 (C4.5).....	77
Tabel 4.38	K 10 <i>Fold Cross Validation</i> 70:30 (C4.5).....	78
Tabel 4.39	K 10 <i>Fold Cross Validation</i> 60:40 (C4.5).....	79

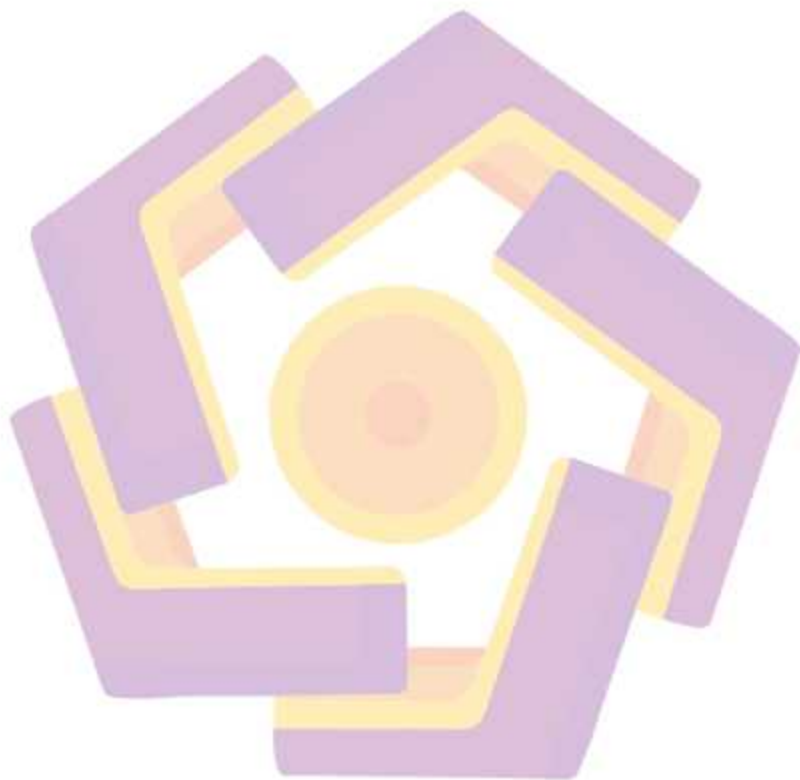
Tabel 4.40 Performa data uji (C4.5).....	79
Tabel 4.41 Nilai ROC (AUC) data uji (C4.5).....	80
Tabel 4.42 <i>Classification Report</i> 90:10 data uji (C4.5).....	82
Tabel 4.43 <i>Classification Report</i> 80:20 data uji (C4.5).....	82
Tabel 4.44 <i>Classification Report</i> 70:30 data uji (C4.5).....	83
Tabel 4.45 <i>Classification Report</i> 60:40 data uji (C4.5).....	83
Tabel 4.46 Jangkauan Parameter GridSearchCV (C4.5).....	83
Tabel 4.47 Hasil parameter Terbaik GridSearchCV (C4.5).....	84
Tabel 4.48 Hasil GridSearchCV data latih dan uji (C4.5).....	84
Tabel 4.49 GridSearchCV K 10 <i>Fold Cross Validation</i> 90:10 (C4.5).....	85
Tabel 4.50 GridSearchCV K 10 <i>Fold Cross Validation</i> 80:20 (C4.5).....	86
Tabel 4.51 GridSearchCV K 10 <i>Fold Cross Validation</i> 70:30 (C4.5).....	86
Tabel 4.52 GridSearchCV K 10 <i>Fold Cross Validation</i> 60:40 (C4.5).....	87
Tabel 4.53 Performa GridSearchCV data uji (C4.5).....	87
Tabel 4.54 Nilai ROC (AUC) GridSearchCV data uji (C4.5).....	88
Tabel 4.55 <i>Classification Report</i> GridSearchCV 90:10 data uji (C4.5).....	90
Tabel 4.56 <i>Classification Report</i> GridSearchCV 80:20 data uji (C4.5).....	90
Tabel 4.57 <i>Classification Report</i> GridSearchCV 70:30 data uji (C4.5).....	90
Tabel 4.58 <i>Classification Report</i> GridSearchCV 60:40 data uji (C4.5).....	91
Tabel 4.59 Hasil Evaluasi (XGBoost).....	91
Tabel 4.60 Hasil Evaluasi (C4.5).....	92
Tabel 4.61 Hasil Evaluasi GridSearchCV 10 <i>Fold</i> (XGBoost).....	92
Tabel 4.62 Hasil Evaluasi GridSearchCV 10 <i>Fold</i> (C4.5).....	93
Tabel 4.63 Perbandingan <i>Confusion Matrix</i> (XGBoost).....	94
Tabel 4.64 Perbandingan <i>Confusion Matrix</i> (C4.5).....	94
Tabel 4.65 Perbandingan <i>Confusion Matrix</i> GridSearchCV (XGBoost).....	95
Tabel 4.66 Perbandingan <i>Confusion Matrix</i> GridSearchCV (C4.5).....	95

DAFTAR GAMBAR

Gambar 2.1	Jenis-jenis Algoritma Data Mining.....	18
Gambar 2.2	Proses <i>Knowledge Discovery Database</i> (KDD).....	20
Gambar 2.3	Contoh Pohon Keputusan.....	22
Gambar 2.4	Proses Penyelesaian Algoritma C4.5.....	24
Gambar 2.5	Confusion Matrix.....	28
Gambar 2.6	Proses dari K-Fold.....	31
Gambar 3.1	Alur Penelitian.....	34
Gambar 3.2	Dataset Applicant Details di situs Kaggle.....	35
Gambar 3.3	Data <i>validation</i> (Cek nilai null).....	36
Gambar 3.4	Data <i>validation</i> (Cek duplikat).....	36
Gambar 3.5	Data <i>cleaning</i> (Ubah nama atribut).....	37
Gambar 3.6	Data <i>cleaning</i> (Penghapusan Atribut).....	37
Gambar 3.7	<i>Scaling</i> (Konversi).....	38
Gambar 3.8	Label encoding.....	38
Gambar 3.9	<i>Oversampling</i>	39
Gambar 3.10	<i>Correlation</i> /Hubungan atribut.....	39
Gambar 3.11	Split Data.....	41
Gambar 3.12	Model biasa (XGBoost).....	42
Gambar 3.13	Model biasa (C4.5).....	42
Gambar 3.14	10 <i>K Fold Cross Validation</i> (XGBoost).....	43
Gambar 3.15	10 <i>K Fold Cross Validation</i> (C4.5).....	44
Gambar 3.16	GridSearchCV (XGBoost).....	45
Gambar 3.17	GridSearchCV (C4.5).....	46
Gambar 3.18	Mean 10 <i>K Fold Cross Validation</i> (Evaluasi).....	47
Gambar 3.19	Data uji (Evaluasi).....	48
Gambar 3.20	ROC (AUC) (Evaluasi).....	48
Gambar 3.21	Confusion Matrix (Evaluasi).....	49
Gambar 3.22	Classification report (Evaluasi).....	49
Gambar 3.23	Dataset Applicant Details.....	51
Gambar 3.24	Dataset Applicant Details lanjutan.....	51
Gambar 4.1	Hasil <i>Correlation</i> (Hubungan antar atribut).....	57
Gambar 4.2	Confusion Matrix 90:10 dan 80:20 data uji (XGBoost).....	64
Gambar 4.3	Confusion Matrix 70:30 dan 60:40 data uji (XGBoost).....	64
Gambar 4.4	Confusion Matrix GridSearchCV 90:10 dan 80:20 (XGBoost).....	73
Gambar 4.5	Confusion Matrix GridSearchCV 70:30 dan 60:40 (XGBoost).....	73
Gambar 4.6	Confusion Matrix 90:10 dan 80:20 data uji (C4.5).....	81
Gambar 4.7	Confusion Matrix 70:30 dan 60:40 data uji (C4.5).....	81
Gambar 4.8	Confusion Matrix GridSearchCV 90:10 dan 80:20 data uji (C4.5).....	89
Gambar 4.9	Confusion Matrix GridSearchCV 70:30 dan 60:40 data uji (C4.5).....	89
Gambar 4.10	Tampilan <i>prototype</i> model.....	97
Gambar 4.11	Tampilan <i>prototype</i> model Lanjutan.....	97

DAFTAR LAMPIRAN

Lampiran 1 Tampilan Web Publikasi buat Jurnal Skripsi.....	105
--	-----



INTISARI

Penelitian ini membandingkan algoritma XGBoost dan C4.5 dalam memprediksi dengan cara mengklasifikasikan risiko kredit lebih baik berdasarkan data pemohon untuk menentukan kelayakan pinjaman. Masalah utama adalah identifikasi faktor-faktor yang mempengaruhi risiko kredit dan penggunaan algoritma yang tepat untuk meningkatkan akurasi prediksi. Risiko kredit yang tidak teridentifikasi dapat mengakibatkan kerugian bagi lembaga keuangan dan peminjam.

Metode penelitian menggunakan pendekatan Knowledge Discovery in Data (KDD), mencakup seleksi, pra-pemrosesan, dan transformasi data, serta pembentukan model untuk mengklasifikasikan risiko kredit dengan XGBoost dan C4.5. Data diproses menggunakan K Fold cross-validation 10 lipatan dengan rasio 90:10, 80:20, 70:30, dan 60:40. Analisis dilakukan dengan Python dan Google Colab, serta optimasi model menggunakan GridSearchCV. Kinerja algoritma diuji menggunakan metrik akurasi, presisi, recall, F1-score, dan confusion matrix.

Berdasarkan hasil perbandingan kinerja algoritma C4.5 dan XGBoost, dapat disimpulkan bahwa algoritma C4.5 menunjukkan performa lebih unggul dalam memprediksi risiko kredit, baik sebelum maupun setelah penerapan optimasi dengan GridSearchCV. C4.5 mencapai akurasi tertinggi 96.20% dan nilai ROC (AUC) 0.9730 di rasio data 80:20. Meskipun XGBoost menunjukkan peningkatan performa yang hampir sama dengan C4.5 setelah optimasi, C4.5 tetap unggul dalam hal akurasi sampai efisiensi waktu. Oleh karena itu, C4.5 direkomendasikan untuk klasifikasi risiko kredit, sementara XGBoost tetap menjadi alternatif kompetitif, terutama dengan optimasi parameter.

Kata kunci: XGBoost, C4.5, risiko kredit, kelayakan pinjaman, machine learning.

ABSTRACT

This study compares the XGBoost and C4.5 algorithms in predicting and classifying credit risk based on borrower data to determine loan eligibility. The main issue is identifying factors that affect credit risk and using the right algorithms to improve prediction accuracy. Unidentified credit risks can result in losses for financial institutions and borrowers.

The research method uses a Knowledge Discovery in Data (KDD) approach, including data selection, preprocessing, and transformation, as well as model formation for credit risk classification with XGBoost and C4.5. Data processing is performed using 10-fold cross-validation with ratios of 90:10, 80:20, 70:30, and 60:40. The analysis is conducted using Python and Google Colab, with model optimization through GridSearchCV. Algorithm performance is evaluated using accuracy, precision, recall, F1-score, and confusion matrix metrics.

Based on the performance comparison, it can be concluded that the C4.5 algorithm outperforms XGBoost in predicting credit risk, both before and after optimization with GridSearchCV. C4.5 achieves the highest accuracy of 96.20% and ROC (AUC) value of 0.9730 data ratio 80:20. Although XGBoost shows performance improvements similar to C4.5 after optimization, C4.5 remains superior in accuracy to time efficiency. Therefore, C4.5 is recommended for credit risk classification, while XGBoost remains a competitive alternative, especially with parameter optimization.

Keyword: XGBoost, C4.5, credit risk, loan eligibility, machine learning