

**PENILAIAN KREDIT MENGGUNAKAN ALGORITMA
XGBOOST DAN LOGISTIC REGRESSION**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



diajukan oleh
GANI RAMADHANI
18.11.2037

Kepada
**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**PENILAIAN KREDIT MENGGUNAKAN ALGORITMA
XGBOOST DAN LOGISTIC REGRESSION**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



diajukan oleh

GANI RAMADHANI

18.11.2037

Kepada

PROGRAM SARJANA

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2022

HALAMAN PERSETUJUAN

SKRIPSI

**PENILAIAN KREDIT MENGGUNAKAN ALGORITMA
XGBOOST DAN LOGISTIC REGRESSION**

yang disusun dan diajukan oleh

Gani Ramadhan

18.11.2037

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 17 Oktober 2022

Dosen Pembimbing

Ainul Yaqin, M. Kom

NIK. 190302255

HALAMAN PENGESAHAN

SKRIPSI

**PENILAIAN KREDIT MENGGUNAKAN ALGORITMA
XGBOOST DAN LOGISTIC REGRESSION**

yang disusun dan diajukan oleh

Gani Ramadhan!

18.11.2037

Telah dipertahankan di depan Dewan Pengüji
pada tanggal 17 Oktober 2022

Susunan Dewan Pengüji

Nama Pengüji

Tanda Tangan

Mardhiya Hayaty, S.T., M.Kom
NIK. 190302108

Anna Balta, M.Kom
NIK. 190302290

Alnul Yaqin, M. Kom
NIK. 190302255

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 17 Oktober 2022

DEKAN FAKULTAS ILMU KOMPUTER

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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Gani Ramadhani
NIM : 18.11.2037

Menyatakan bahwa Skripsi dengan judul berikut:

PENILAIAN KREDIT MENGGUNAKAN ALGORITMA XGBOOST DAN LOGISITIC REGRESSION

Dosen Pembimbing : Ainul Yaqin, 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, 17 Oktober 2022

Yang Menyatakan,



Gani Ramadhani

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah SWT atas kasih sayang dan rahmat-Nya sekaligus telah memberikan kesehatan dan kemampuan sehingga dapat menyelesaikan penelitian ini. Tidak lupa pula shalawat dan salam senantiasa tercurahkan bagi Rasulullah SAW yang telah membawa manusia dari zaman kegelapan menuju zaman terang benderang.

Penelitian sekaligus penulisan naskah skripsi ini tidak dapat selesai tanpa dukungan dan bantuan dari berbagai pihak. Saya ucapkan banyak terima kasih kepada :

1. Kedua orang tua saya yang telah membiayai perkuliahan dan memberikan dukungan serta doa sehingga bisa menyelesaikan penelitian ini.
2. Bapak Ainul Yaqin, M.Kom. selaku dosen pembimbing yang telah memberikan arahan sekaligus bimbingan dalam penulisan skripsi ini.
3. Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak bisa penulis sebutkan semuanya.

Yogyakarta, 18 November 2022

Gani Ramadhani

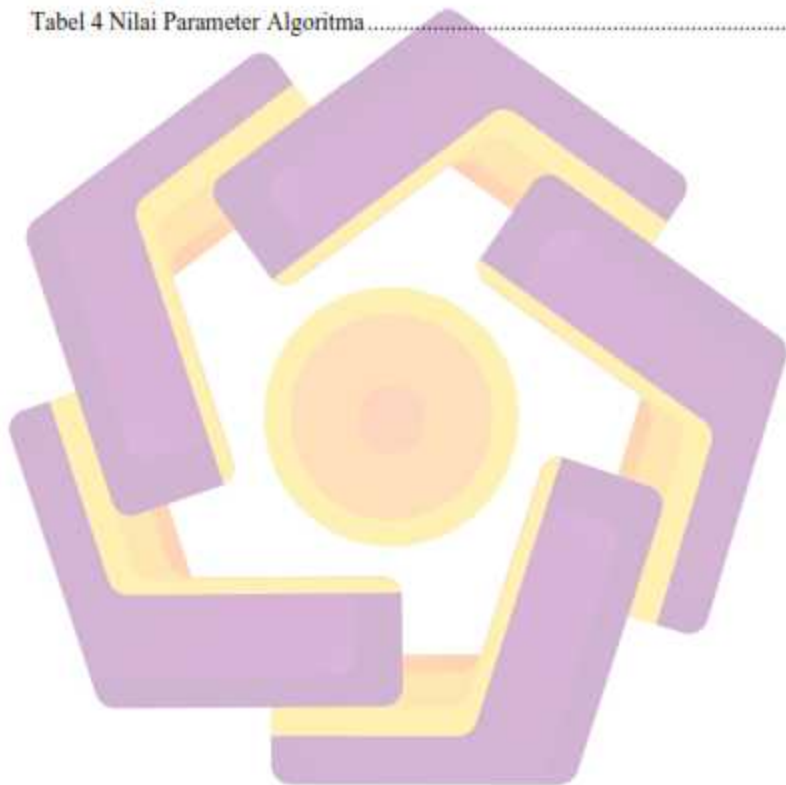
DAFTAR ISI

HALAMAN JUDUL	2
DAFTAR ISI	5
DAFTAR TABEL	9
DAFTAR GAMBAR	10
DAFTAR ISTILAH	11
INTISARI	12
Abstract	13
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan masalah	2
1.3 Tujuan Penelitian	2
1.4 Batasan Masalah	2
1.5 Manfaat Penelitian	2
BAB II TINJAUAN PUSTAKA	3
2.1 Literature Review	3
2.2 Landasan Teori	5
2.2.1 Penilaian Kredit	5
2.2.2 Stratified K-Fold Cross Validation	5
2.2.3 XGBoost	6
2.2.4 Logistic Regression	9
2.2.5 Confusion Matrix	10
BAB III METODOLOGI PENELITIAN	12
3.1 Analisis Kebutuhan Penelitian	12
3.1.1 Perangkat Keras	12
3.1.2 Perangkat Lunak	12
3.1.2 Data Penelitian	12
3.2 Alur Penelitian	14
3.2.1 Pengambilan Dataset	15
3.2.2 Data Preprocessing	15
3.2.1 Pembagian Dataset	17

3.2.2 Training Model	17
3.2.3 Testing Model	18
3.2.4 Evaluasi Model	18
BAB IV HASIL DAN PEMBAHASAN	19
4.1 Implementasi	19
4.1.1 Import Library	19
4.1.2 Konversi Dataset Menjadi Dataframe dan Menambahkan Kolom	20
4.1.3 Feature Selection	21
4.1.4 Memisahkan Kolom Atribut dan Label	22
4.1.5 Pembuatan Model	23
4.1.6 Normalisasi, Split, Training Sekaligus Testing Model	24
4.1.7 Konversi Hasil Testing Menjadi Dataframe	26
4.2 Hasil Pengujian	28
BAB V KESIMPULAN DAN SARAN	33
5.1 Kesimpulan	33
5.2 Saran	33
DAFTAR PUSTAKA	34

DAFTAR TABEL


Tabel 1 Perbandingan Penelitian Penilaian Kredit	4
Tabel 2 Penelitian Menggunakan Xgboost dan Logistic Regression	4
Tabel 3 Sample Dataset	13
Tabel 4 Nilai Parameter Algoritma	18



DAFTAR GAMBAR

Gambar 1 Ilustrasi Stratified K-Fold Cross Validation	6
Gambar 2 Ilustrasi XGBoost	8
Gambar 3 Ilustrasi Fungsi Logit	9
Gambar 4 Ilustrasi Fungsi Sigmoid	10
Gambar 5 Ilustrasi Confusion Matrix	11
Gambar 6 Proporsi Kelas Dataset	13
Gambar 7 Alur Penelitian	14
Gambar 8 Setelah Dilakukan Data <i>Scaling</i>	16
Gambar 9 Korelasi Variabel Input dan Output	17
Gambar 10 Sample Dataset Format DAT	21
Gambar 11 Sampel <i>Dataframe</i>	21
Gambar 12 Hasil <i>Feature Selection</i>	22
Gambar 13 Data Atribut	23
Gambar 14 Data Label	23
Gambar 15 Hasil Algoritma XGBoost	25
Gambar 16 Hasil Algoritma Logistic Regression	25
Gambar 17 <i>Dataframe</i> Hasil Akhir	27
Gambar 18 Perbandingan <i>Accuracy</i>	28
Gambar 19 Perbandingan <i>Precision</i>	29
Gambar 20 Perbandingan <i>Recall</i>	30
Gambar 21 Perbandingan <i>F1 Score</i>	31
Gambar 22 Perbandingan Seluruh Metode	32

DAFTAR ISTILAH



Library	sekumpulan kode yang dapat dipanggil kedalam program lain.
Residual	selisih antara nilai asli dengan nilai prediksi.
Node	sebuah struktur pohon yang mengandung nilai atau kondisi.
Learning rate	parameter untuk menghitung nilai koreksi bobot pada proses <i>training</i> .
Matriks	susunan bilangan dalam kolom dan baris.
True Positive	prediksi positif dan benar.
False Positive	prediksi positif namun salah.
True Negative	prediksi negatif dan benar.
False Negative	prediksi negatif namun salah.
Dictionary	tipe data pada <i>python</i> untuk menyimpan kumpulan data dengan struktur <i>key</i> dan <i>value</i> .
Hyperparameter Tunning	proses menemukan nilai parameter terbaik.

INTISARI

Penilaian kredit merupakan suatu proses atau sistem yang digunakan oleh lembaga pembiayaan atau bank untuk menilai kelayakan seseorang yang mengajukan pinjaman. Hal ini sangat diperlukan untuk menghindari kerugian akibat gagal bayar. Menanggapi hal tersebut dibutuhkan sebuah metode yang efisien, cepat dan akurat untuk mengklasifikasikan layak atau tidaknya seseorang untuk diberikan pinjaman.

Penulis mengusulkan metode *machine learning* dan membandingkan algoritma *XGBoost* dengan *logistic regression*. Penelitian ini menggunakan dataset penilaian kredit Australia yang didapat dari situs *UCI Machine Learning Repository*. Dilakukan beberapa proses *pre-processing* pada dataset yaitu normalisasi dan *feature selection*.

Setelah dilatih dan diuji dengan *stratified kfold cross validation*, *XGBoost* menghasilkan rata-rata akurasi 85,51%; *F1 Score* 83,81%; *precision* 83,80% dan *recall* 84,04% sedangkan *logistic regression* menghasilkan rata-rata akurasi 85,94%; *F1 Score* 85,36%; *precision* 80,08%; dan *recall* 91,52%. Kedua algoritma dapat mengklasifikasikan layak atau tidaknya seseorang untuk diberikan pinjaman dengan baik, sehingga dapat digunakan untuk membantu institusi keuangan maupun para analis kredit.

Kata kunci: penilaian kredit, klasifikasi, *XGBoost*, *logistic regression*, *data mining*

Abstract

Credit Scoring is a process or system used by financial institutions or banks to assess the eligibility of a person applying for a loan. This is very necessary to avoid losses due to default. Responding to this, an efficient, fast and accurate method is needed to classify whether or not someone is eligible for a loan.

The author proposes a machine learning method and compares the XGBoost algorithm with logistic regression. This research uses the Australian credit scoring dataset obtained from the UCI Machine Learning Repository. Some pre-processing was carried out on the dataset, namely normalization and feature selection.

After being trained and tested with stratified kfold cross validation, XGBoost produces an average accuracy of 85,51%; F1 Score 83,81%; precision 83.80% and recall 84,04% while logistic regression produces an average accuracy of 85,94%; F1 Score 85,36%; precision 80,08%; and 91,52% recalls. Both algorithms can classify whether or not a person is eligible for a loan properly, so that it can be used to help financial institutions and credit analysts.

Keyword: credit scoring, classification, XGBoost, logistic regression, data mining