

**ANALISIS PERBANDINGAN ALGORITMA LOGISTIC REGRESSION,
K- NEAREST NEIGHTBORS DAN RANDOM FOREST DALAM PREDIKSI
PINJAM ONLINE DENGAN METODE CRIPS-DM
SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh
SAUSAN QOTHRUNNADA
21.12.2198

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**ANALISIS PERBANDINGAN ALGORITMA LOGISTIC REGRESSION,
K- NEAREST NEIGHTBORS DAN RANDOM FOREST DALAM PREDIKSI
PINJAM ONLINE DENGAN METODE CRIPS-DM**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Sistem Informasi



disusun oleh

SAUSAN QOTHRUNNADA

21.12.2198

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS PERBANDINGAN ALGORITMA LOGISTIC REGRESSION,
K-NEAREST NEIGHBORS DAN RANDOM FOREST DALAM PREDIKSI
PINJAM ONLINE DENGAN METODE CRIPS-DM**

yang disusun dan diajukan oleh

SAUSAN QOTHRUNNADA

21.12.2198

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 Mei 2025

Dosen Pembimbing,


Anggit Dwi Hartanto, S.Kom., M.Kom
NIK. 190302163

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS PERBANDINGAN ALGORITMA LOGISTIC REGRESSION,
K- NEAREST NEIGHBORS DAN RANDOM FOREST DALAM PREDIKSI
PINJAM ONLINE DENGAN METODE CRIPS-DM**

yang disusun dan diajukan oleh

SAUSAN QOTHRUNNADA

21.12.2198

Telah dipertahankan di depan Dewan Penguji
pada tanggal 28 Mei 2025

Susunan Dewan Penguji

Nama Penguji

Yoga Pristyanto, S.Kom., M.Eng
NIK. 190302412

Ika Nur Fajri, S.Kom., M.Kom
NIK. 190302268

Anggit Dwi Hartanto, S.Kom., M.Kom
NIK. 190302163

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 28 Mei 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Sausan Qothrunnada
NIM : 21.12.2198

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Perbandingan Algoritma Logistic Regression, KNN, dan Random Forest dalam
Prediksi Pinjaman Online dengan Metode CRISP-DM

Dosen Pembimbing : Anggit Dwi Hartanto, 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, 28 Mei 2025

Yang Menyatakan,


Sausan Qothrunnada

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur dan kerendahan hati kepada Allah SWT, Tuhan Yang Maha Esa atas limpahan rahmat, hidayah, dan kekuatan yang telah diberikan sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Karya ini juga penulis persembahkan kepada kedua orang tua tercinta dan terkasih bapak Untung Sugiri dan ibu Fatonah, yang senantiasa memberikan cinta, dukungan, semangat, dan do'a tanpa henti dalam setiap langkah serta pengorbanan yang tak pernah terucapkan namun selalu terasa dalam setiap kehidupan penulis, because you are the main reason I stand and keep walking in this life.

Ucapan terima kasih penulis sampaikan kepada dosen pembimbing bapak Anggit Dwi Hartanto, M.Kom yang telah memberikan arahan dan bimbingan, serta ilmu yang bermanfaat selama proses penyusunan skripsi ini. Tidak lupa, penulis juga mempersembahkan karya ini kepada teman teman sepejuangan SI06 dan sahabat – sahabat tercinta Luh Ajeng Roro Mangli, Sintia Eka Purwati, Maysi Margianti, Syahnaz Tasya Nazwa Idwal dan Ibnu Muhammad yang selalu memberikan semangat, kebersamaan, serta dorongan moral selama proses perjuangan ini berlangsung.

Ucapan terima kasih juga penulis sampaikan kepada diri sendiri yang telah bertahan selama ini. Walaupun berkali kali rasanya ingin menyerah saat semuanya terasa sulit tetapi memilih tetap berjalan, meskipun jalan yang dilalui terkadang sunyi, terjal, dan melelahkan. Terima kasih karena sudah bertahan dengan keadaan tubuh yang tidak selalu kuat, untuk hari - hari yang dilewati dengan rasa sakit ditubuh yang selalu dihiraukan setiap kali sakit ini muncul. Terimakasih tetap bertahan dengan semua keraguan yang ada namun tak pernah berhenti berharap.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa, yang telah melimpahkan rahmat, hidayah, serta karunia-Nya, sehingga penulis dapat menyelesaikan laporan tugas akhir ini dengan lancar. Tugas akhir ini disusun untuk memenuhi salah satu syarat kelulusan Program Sarjana pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Selain itu penulis dengan segala kerendahan hati ingin menyampaikan rasa terima kasih kepada semua pihak yang telah berjasa memberikan dukungan dan bantuan untuk menyelesaikan laporan tugas akhir ini. Untuk itu penulis mengucapkan terima kasih kepada :

1. Prof. Dr. M Suyanto, MM. selaku Rektor Universitas Amikom Yogyakarta
2. Hanif al fatta, S.Kom., M.Kom.Ph.D selaku Dekan Program Fakultas Ilmu Komputer
3. Anggit Dwi Hartanto, M.Kom selaku Ketua Program Studi Sistem Informasi
4. Anggit Dwi Hartanto, M.Kom selaku dosen pembimbing yang memberikan arahan, saran, dan motivasi terhadap penulis
5. Kedua orang tua bapak Untung Sugiri, ibu Fatonah, kakak Nisa Agustin, & adik Fayyadh Raffay Rabbani yang senantiasa memberikan cinta, dukungan, semangat, dan do'a tanpa henti dalam setiap Langkah.
6. Sahabat seperjuangan, Ibnu Muhammad, Luh Ajeng Roro Mangli, Sintia Eka Purwati, Maysi Margianti, dan Syahnaz Tasya Nazwa Idwal yang selalu memberikan semangat, kebersamaan, serta dorongan moral selama proses perjuangan ini berlangsung.
7. Teman-teman Sistem Informasi 06 angkatan 21 atas kebersamaan dan Kerjasamanya selama dibangku perkuliahan ini.

Yogyakarta, 18 Mei 2025



Sausan Qothrunnada

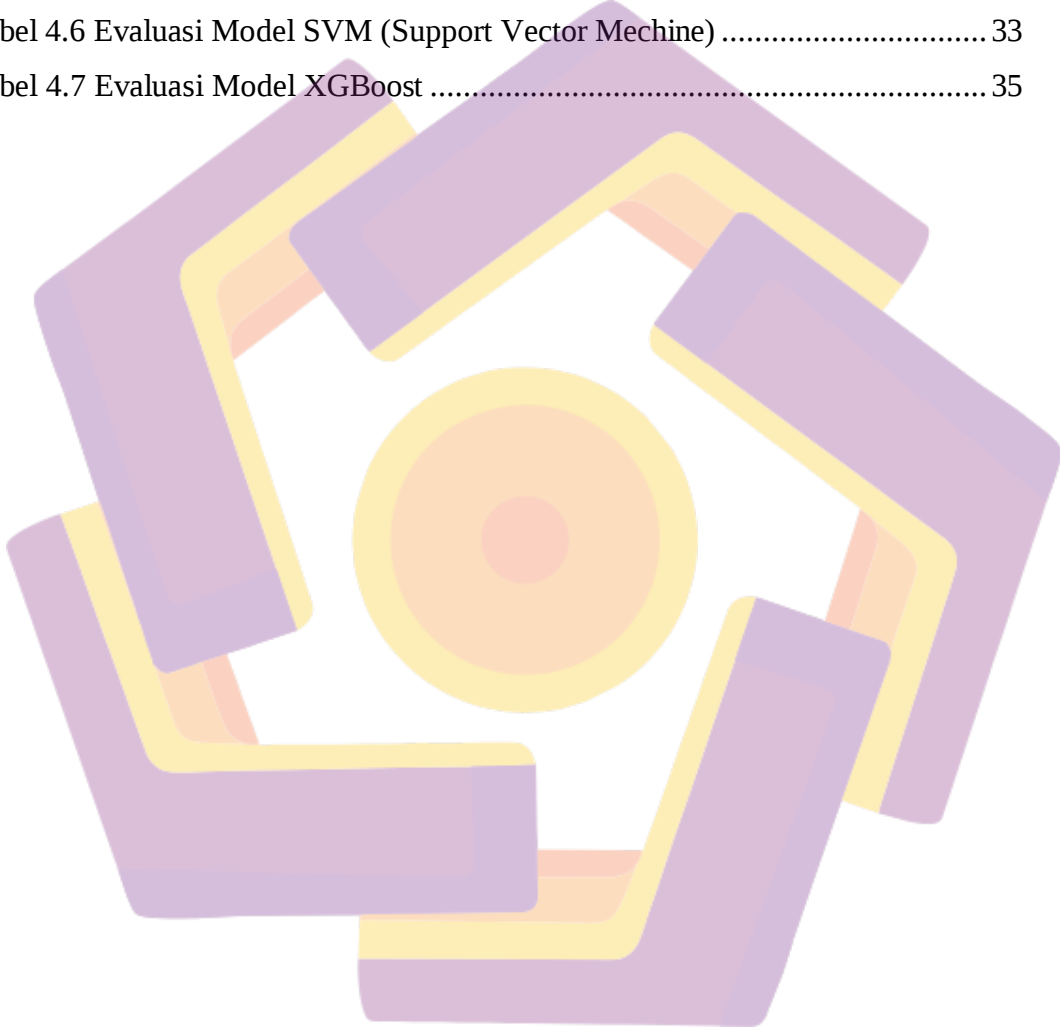
DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN SKRIPSI.....	ii
HALAMAN PENGESAHAN SKRIPSI	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMBANG DAN SINGKATAN.....	xi
DAFTAR ISTILAH.....	xii
INTISARI	xiv
ABSTRACT.....	xv
BAB I.....	1
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	3
1.6 Sistematika Penulisan.....	4
BAB II.....	6
TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	13
2.2.1 Pinjaman Online	13
2.2.2 Machine Learning	13
2.2.3 Logistic Regression	13
2.2.5 Random Forest	14
2.2.6 Crips-DM.....	14

2.2.7 Google Collaboratory	15
2.2.8 Phyton.....	15
2.2.9 Data Mining	15
BAB III	16
METODE PENELITIAN.....	16
3.1 Objek Penelitian.....	16
3.2.1 Model Proses CRIPS-DM.....	17
3.3 Alat dan Bahan.....	22
3.4 Data Penelitian	22
BAB IV	23
HASIL DAN PEMBAHASAN	23
4.1 Identifikasi Masalah.....	23
4.2 Pengumpulan Data	23
4.3 CRIPS-DM	23
4.3.1 Bussines Understanding.....	23
4.3.2 Understanding	24
4.3.3 Data Preparation.....	25
4.3.4 Modeling.....	27
4.3.5 Evaluasi.....	38
4.3.6 Deployment.....	39
BAB V.....	42
PENUTUP	42
5.1 Kesimpulan	42
5.2 Saran.....	42
REFERENSI.....	44

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	9
Tabel 4.1 Proses Tranformasi	26
Tabel 4.2 Proses Konversi Data.....	26
Tabel 4.3 Evaluasi Model Logistic Regression	27
Tabel 4.4 Evaluasi Model KNN.....	29
Tabel 4.5 Evaluasi Model Random Forest	31
Tabel 4.6 Evaluasi Model SVM (Support Vector Mechine)	33
Tabel 4.7 Evaluasi Model XGBoost	35

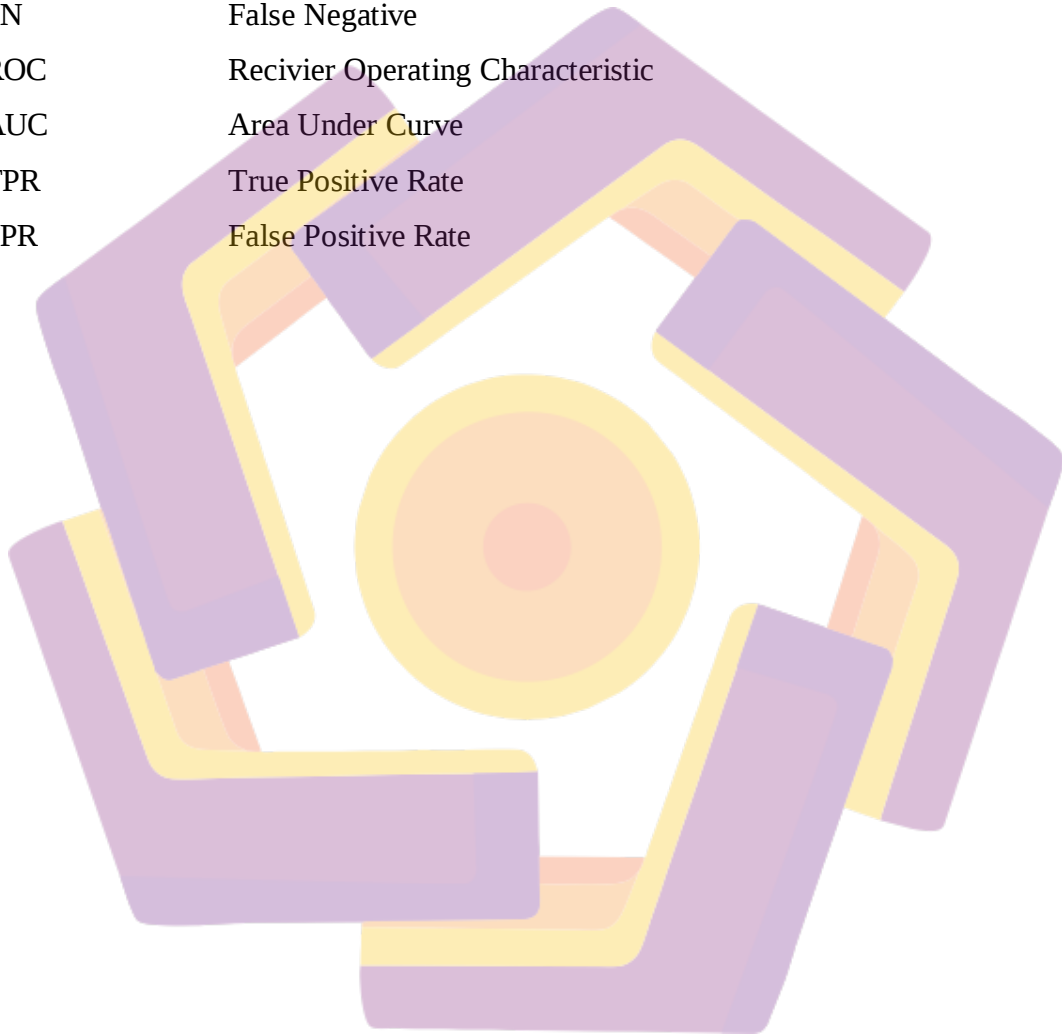


DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	17
Gambar 4.1 Data Understanding.....	25
Gambar 4.2 Missing Value	27
Gambar 4.3 Outpu Visual Heatmap Confusion Matrik.....	28
Gambar 4.4 Outpu Visual Heatmap Clasification.....	28
Gambar 4.5 Outpu Visual Heatmap Confusion Matrik.....	30
Gambar 4.6 Outpu Visual Heatmap classificaton Matrik.....	31
Gambar 4.7 Outpu Visual Heatmap Confusion Matrik.....	32
Gambar 4.8 Outpu Visual Heatmap Classification Matrik.....	33
Gambar 4.9 Outpu Visual Heatmap Confusion Matrik.....	34
Gambar 4.10 Outpu Visual Heatmap Classification Matrik.....	35
Gambar 4.11 Outpu Visual Heatmap Confusion Matrik.....	36
Gambar 4.12 Outpu Visual Heatmap Classification Matrik.....	37
Gambar 4.13 Hasil akurasi Model	38
Gambar 4.14 Output Deployment Looker Studio.....	39

DAFTAR LAMBANG DAN SINGKATAN

CRIPS-DM	Cross-Industry Standard Process for Data Mining
KNN	K-Nearest Neighbors
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
TPR	True Positive Rate
FPR	False Positive Rate

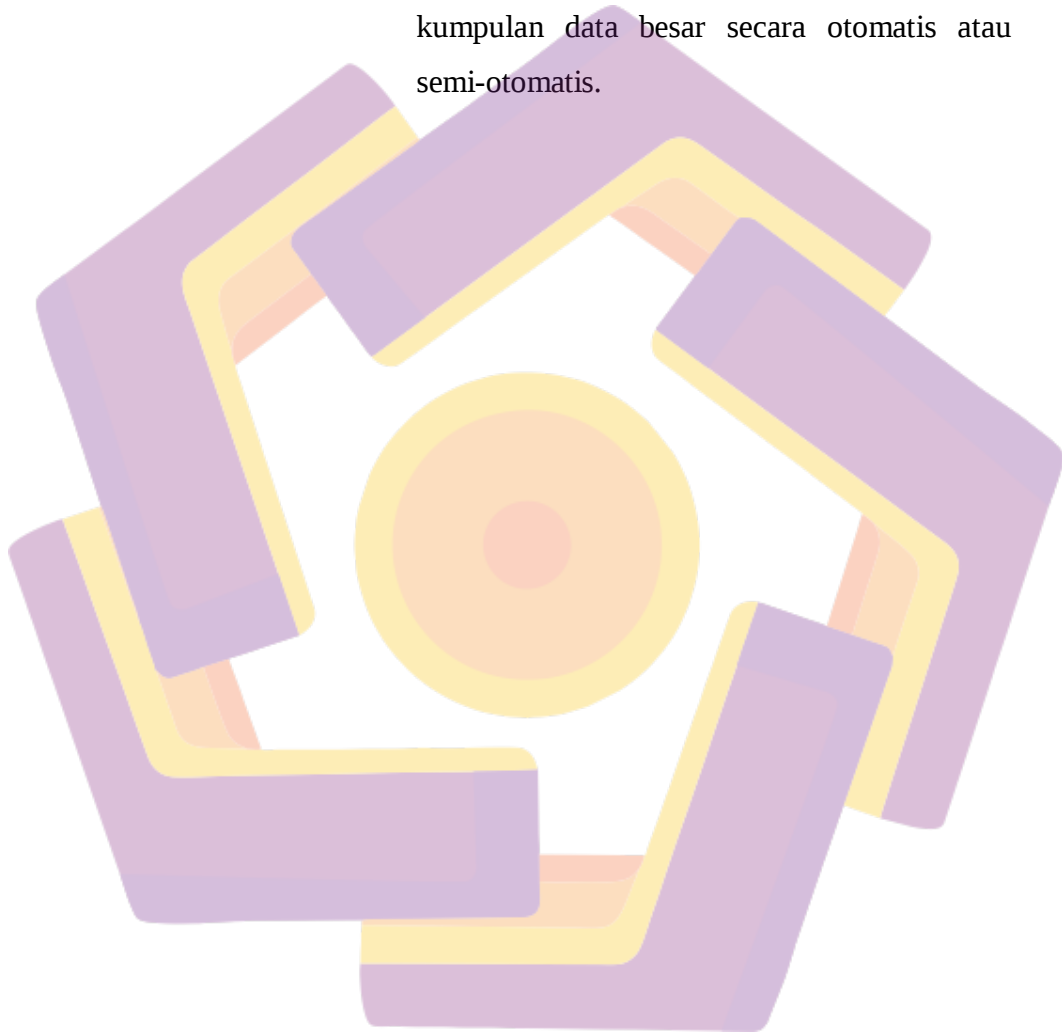


DAFTAR ISTILAH



Pinjaman Online	Layanan pemberian kredit secara digital melalui aplikasi atau platform berbasis internet.
Mechine Learning	Cabang dari kecerdasan buatan yang memungkinkan komputer belajar dari data tanpa diprogram ulang.
Logistic Regression	Algoritma klasifikasi yang memprediksi probabilitas kejadian berdasarkan fungsi logistik
K-Nearest Neighbors (KNN)	Algoritma klasifikasi berbasis jarak, yang menentukan kelas dari objek berdasarkan tetangga terdekatnya
Random Forest	Algoritma ensambel yang membangun banyak pohon keputusan dan menggabungkan hasilnya untuk klasifikasi
Accuracy	Ukuran seberapa sering model membuat prediksi yang benar dari total keseluruhan prediksi.
Precision	Ukuran akurasi prediksi untuk kelas positif dari semua yang diprediksi positif.
Recall (Sensitivity)	Ukuran kemampuan model dalam menemukan seluruh kasus aktual dari kelas positif.
F1-Score	Rata-rata harmonis dari precision dan recall, menggambarkan keseimbangan antara keduanya.
ROC Curve	Grafik untuk mengevaluasi performa klasifikasi dengan membandingkan TPR dan FPR.

AUC (Area Under Curve)	Luas di bawah kurva ROC, menunjukkan seberapa baik model dalam membedakan kelas.
Google Colaboratory (Colab)	Platform pemrograman berbasis cloud yang mendukung eksekusi kode Python untuk keperluan data science.
Data Mining	Proses penggalian informasi atau pola dari kumpulan data besar secara otomatis atau semi-otomatis.



INTISARI

Pesatnya perkembangan teknologi finansial telah mendorong munculnya layanan pinjaman online yang semakin diminati masyarakat karena kemudahan akses, proses cepat, dan minimnya persyaratan. Namun, pertumbuhan ini turut meningkatkan resiko gagal bayar yang dapat merugikan penyedia layanan maupun peminjam. Oleh karena itu, diperlukan sistem prediksi yang andal untuk menilai kelayakan calon peminjam secara objektif, dan akurat guna meminimalisir potensi kerugian.

Penelitian ini bertujuan untuk membandingkan efektivitas lima algoritma Machine Learning, yaitu Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, SVM dan XGBoost dalam memprediksi kelayakan pinjam online. Proses analisis menggunakan pendekatan CRIPS-DM (Cross-Industry Standard Process for Data Mining) yang terdiri dari enam tahap utama : business understanding, data understanding, data preparation, modeling, evaluation, dan Deployment. Dataset yang digunakan berasal dari sumber terbuka Kaggle.com dengan jumlah 4.269 entri data peminjam, yang telah melalui proses pembersihan dan transformasi agar siap dianalisis.

Hasil penelitian menunjukkan bahwa algoritma XGBoost memberikan performa terbaik dengan akurasi 96%, diikuti oleh SVM sebesar 94% dan Logistic Regression sebesar 92%. Pada penelitian ini menunjukkan bahwa XGBoost lebih unggul dalam menangani data yang tidak seimbang dan kompleks. Pada penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem penilaian kredit berbasis data yang lebih akurat dan membantu penyedia layanan pinjam online dalam pengambilan keputusan yang tepat.

Kata kunci: Pinjam online, Machine Learning, Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, SVM, dan XGBoost, CRISP-DM

ABSTRACT

The rapid development of financial technology has encouraged the growth of online loan services, which are increasingly favored by the public due to their ease of access, fast processes, and minimal requirements. However, this growth also increases the risk of default, potentially causing losses for both lenders and borrowers. Therefore, a reliable and objective prediction system is needed to assess borrower eligibility and minimize financial losses.

This study aims to compare the effectiveness of five machine learning algorithms—Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Support Vector Machine (SVM), and XGBoost—in predicting online loan eligibility. The analysis uses the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, which consists of six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The dataset, obtained from Kaggle.com, comprises 4,269 borrower entries that have been cleaned and transformed for analysis.

The results show that the XGBoost algorithm achieves the highest accuracy at 96%, followed by SVM at 94%, and Logistic Regression at 92%. These findings indicate that XGBoost excels in handling imbalanced and complex datasets. This research is expected to contribute to the development of more accurate data-driven credit assessment systems and assist loan providers in making better-informed decisions.

Keywords: *Online loan, Machine Learning, Logistic Regression, K-Nearest Neighbors, Random Forest, SVM, XGBoost, CRISP-DM*