

**DETEKSI KECURANGAN PADA TRANSAKSI E-
COMMERCE MENGGUNAKAN ALGORITMA MACHINE
LEARNING**

SKRIPSI

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



disusun oleh

CHESSA ANINDITYA ANUGRAH PUTRANTO

21.11.4098

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**DETEKSI KECURANGAN PADA TRANSAKSI E-
COMMERCE MENGGUNAKAN ALGORITMA
MACHINE LEARNING**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

CHESSA ANINDITYA ANUGRAH PUTRANTO

21.11.4098

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**DETEKSI KECURANGAN PADA TRANSAKSI E-COMMERCE
MENGUNAKAN ALGORITMA MACHINE LEARNING**

yang disusun dan diajukan oleh

Chessa Aninditya Anugrah Putranto

21.11.4098

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal Kamis, 17 April 2025

Dosen Pembimbing,

Robert Marco, MT, Ph.D
NIK. 190302228

HALAMAN PENGESAHAN
SKRIPSI
DETEKSI KECURANGAN PADA TRANSAKSI E-COMMERCE
MENGGUNAKAN ALGORITMA MACHINE LEARNING

yang disusun dan diajukan oleh

Chessa Aninditya Anugrah Putranto

21.11.4098

Telah dipertahankan di depan Dewan Penguji
pada tanggal Kamis, 17 April 2025

Susunan Dewan Penguji

Nama Penguji

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

Nafiatun Sholihah, S.Kom., M.Cs
NIK. 190302524

Robert Marco, MT, Ph.D
NIK. 190302228

Tanda Tangan

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal Kamis, 17 April 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Chessa Aninditya Anugrah Putranto
NIM : 21.11.4098

Menyatakan bahwa Skripsi dengan judul berikut:

DETEKSI KECURANGAN PADA TRANSAKSI E-COMMERCE MENGUNAKAN ALGORITMA MACHINE LEARNING

Dosen Pembimbing : Robert Marco, MT, 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, Kamis, 17 April 2025

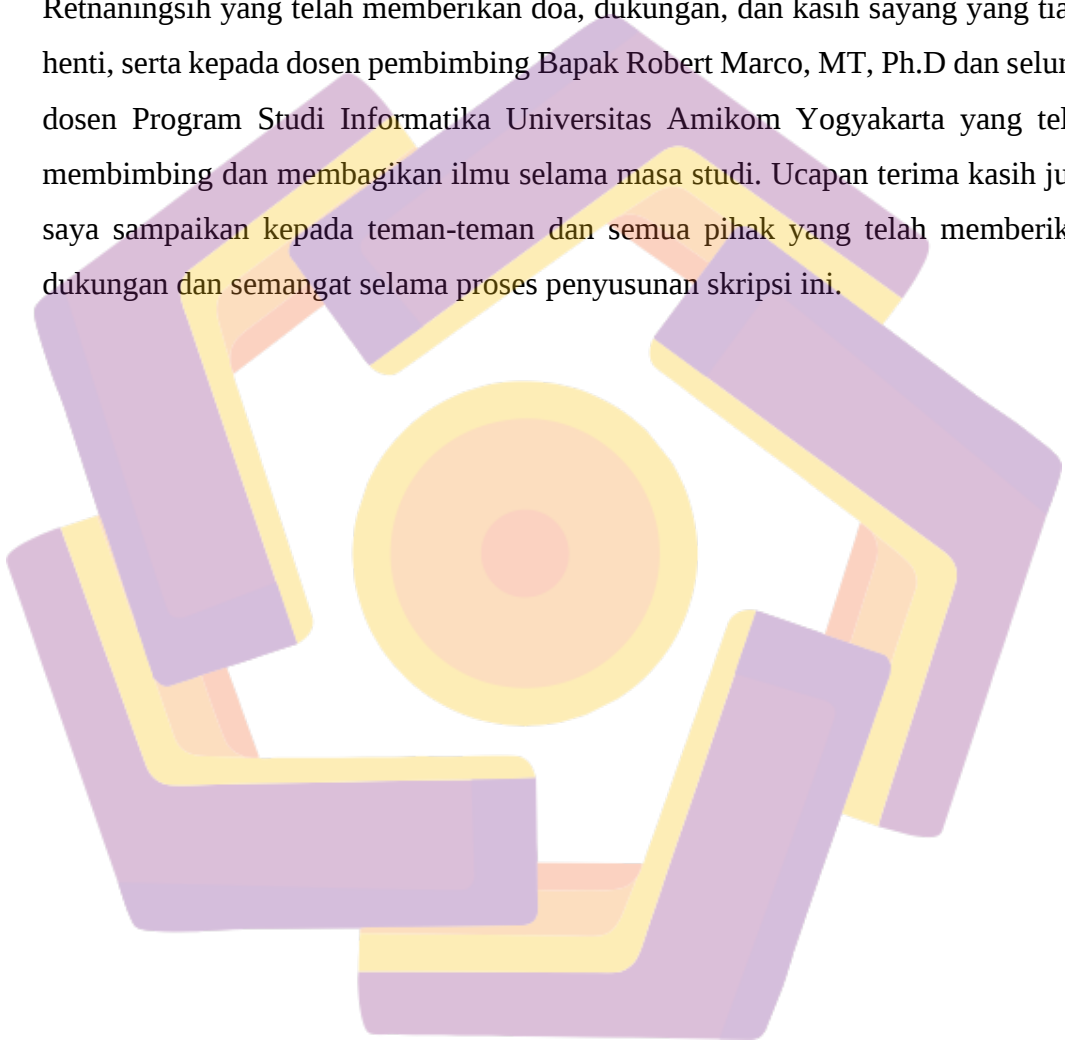
Yang Menyatakan,



Chessa Aninditya Anugrah Putranto

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam, skripsi ini saya persembahkan kepada Allah SWT atas segala rahmat dan hidayah-Nya yang senantiasa menyertai setiap langkah saya. Kepada kedua orang tua tercinta Bapak Dadang Aryanto dan Ibu Endah Tri Retnaningsih yang telah memberikan doa, dukungan, dan kasih sayang yang tiada henti, serta kepada dosen pembimbing Bapak Robert Marco, MT, Ph.D dan seluruh dosen Program Studi Informatika Universitas Amikom Yogyakarta yang telah membimbing dan membagikan ilmu selama masa studi. Ucapan terima kasih juga saya sampaikan kepada teman-teman dan semua pihak yang telah memberikan dukungan dan semangat selama proses penyusunan skripsi ini.



KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT, karena atas limpahan rahmat, karunia, dan izin-Nya, penulis dapat menyelesaikan penyusunan skripsi ini sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, Universitas Amikom Yogyakarta.

Penyusunan skripsi ini tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih dan penghargaan yang setulus-tulusnya kepada:

Bapak Robert Marco, M.T., Ph.D., selaku dosen pembimbing yang telah memberikan waktu, perhatian, arahan, serta bimbingan yang sangat berharga dalam proses penyusunan skripsi ini. Tim dosen penguji yang telah memberikan kritik dan saran yang membangun untuk penyempurnaan karya ini. Seluruh dosen di Program Studi Informatika Universitas Amikom Yogyakarta yang telah memberikan ilmu dan pengalaman selama masa perkuliahan. Kedua orang tua penulis atas doa, dukungan moral dan material, serta kasih sayang yang tiada henti selama proses studi dan penyusunan skripsi. Serta semua pihak yang telah memberikan bantuan dan dukungan dalam bentuk apa pun selama proses penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu, penulis membuka diri terhadap segala bentuk kritik dan saran yang membangun demi perbaikan di masa yang akan datang.

Yogyakarta, 17 April 2025

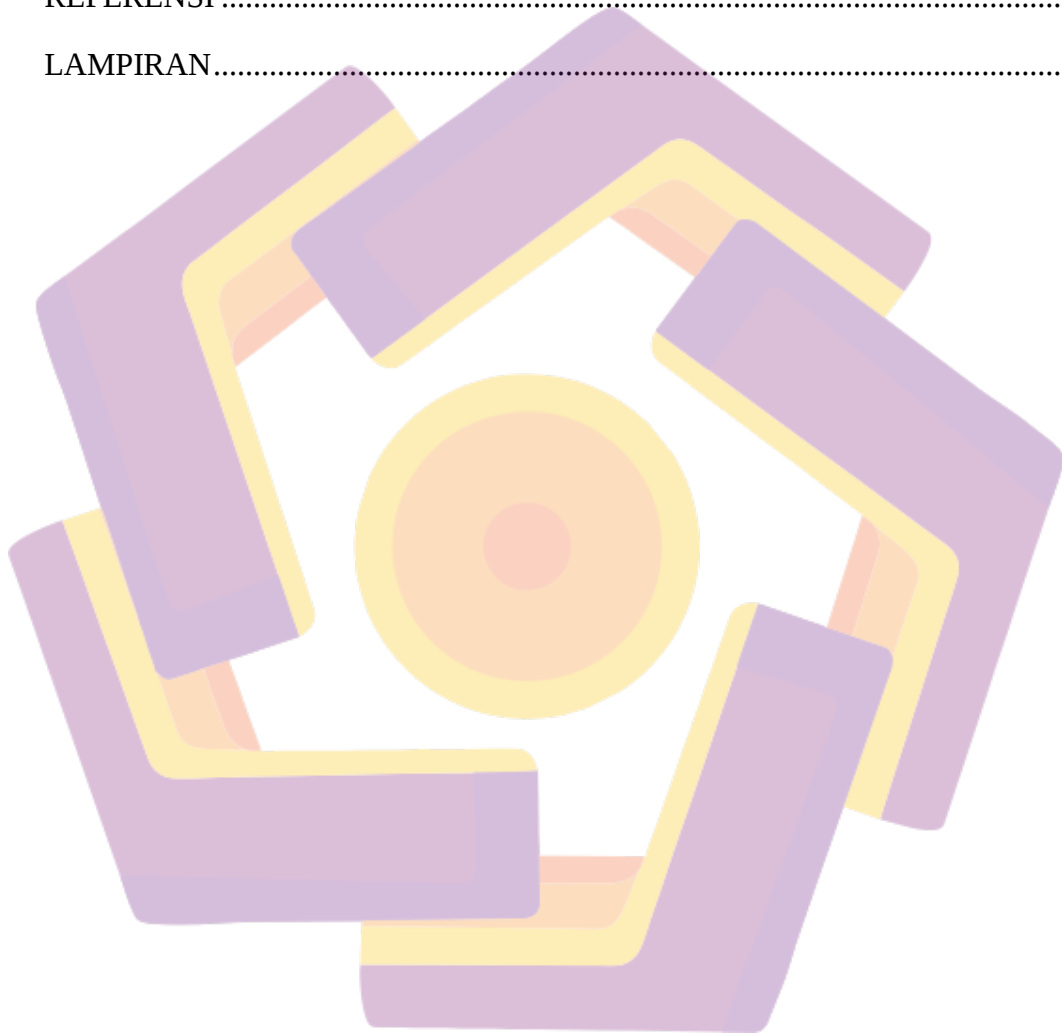
Chessa Aninditya A.P

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.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	6
1.6 Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA	9

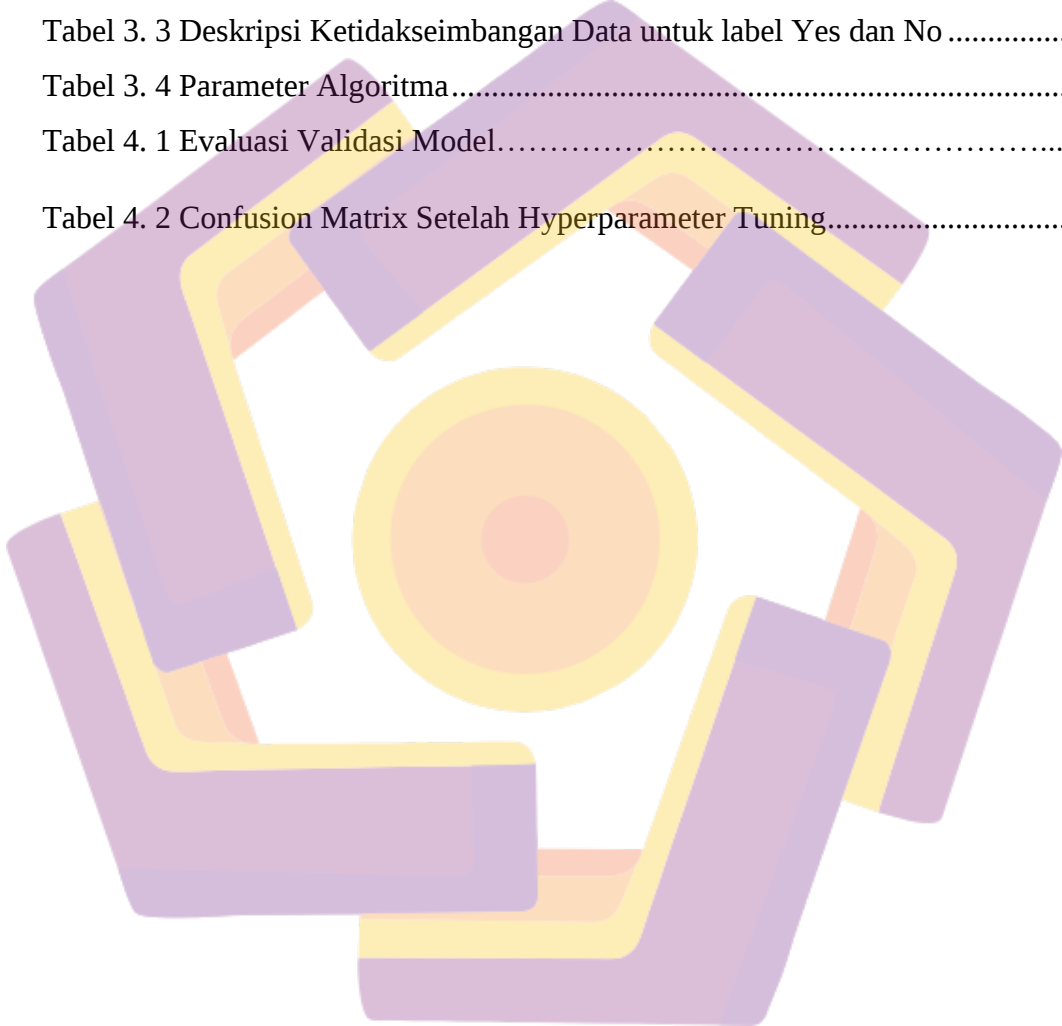
2.1	Studi Literatur	9
2.2	Dasar Teori.....	13
2.2.1	Transaksi E-Commerce.....	13
2.2.2	Ketidakseimbangan Data	13
2.2.3	Metode Resampling	14
2.2.4	Algoritma Random Forest.....	15
2.2.5	Algoritma Extreme Gradient Boost	17
2.2.6	Algoritma Logistic Regression	21
2.2.7	Algoritma Decision Tree.....	25
2.2.8	Algoritma Support Vector Machine (SVM)	26
2.2.9	Algoritma Gaussian Naïve Bayes (GNB)	27
2.2.10	Hyperparameter Tuning Menggunakan Grid Search	28
2.2.11	Evaluasi Performance	30
BAB III METODE PENELITIAN		34
3.1	Alur Penelitian	34
3.2	Alat dan Bahan.....	40
BAB IV HASIL DAN PEMBAHASAN		43
4.1	Proses Pengumpulan Data.....	43
4.2	EDA (Exploratory Data Analysis)	43
4.3	Pre-Processing Data	51
4.3.1	Data Cleaning.....	51
4.3.2	Data Encoding.....	54
4.3.3	Balancing Data.....	55
4.4	Modeling	56
4.6	Hyperparameter Tuning	57

4.7	Evaluasi Model dan Perbandingan Kinerja Algoritma	60
BAB V PENUTUP		69
5.1	Kesimpulan	69
5.2	Saran	69
REFERENSI		71
LAMPIRAN.....		75



DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	12
Tabel 2. 2 <i>Confution Matrix</i>	31
Tabel 3. 1 Deskripsi Fitur Dataset.....	36
Tabel 3. 2 Deskripsi Statistik Data.....	38
Tabel 3. 3 Deskripsi Ketidakseimbangan Data untuk label Yes dan No	38
Tabel 3. 4 Parameter Algoritma.....	40
Tabel 4. 1 Evaluasi Validasi Model.....	61
Tabel 4. 2 Confusion Matrix Setelah Hyperparameter Tuning.....	66



DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	34
Gambar 4. 1 Sampel Dataset.....	43
Gambar 4. 2 Alur Proses EDA.....	44
Gambar 4. 3 Informasi Umum Dataset	45
Gambar 4. 4 Ringkasan Statistik Dataset.....	45
Gambar 4. 5 Visualisasi Fitur Numeric	46
Gambar 4. 6 Visualisasi Fitur Kategorical.....	48
Gambar 4. 7 Informasi jumlah transaksi fraud dan tidak.....	49
Gambar 4. 8 Distribusi usia korban transaksi fraud.....	49
Gambar 4. 9 Perhitungan transaksi fraud dan tidak.....	50
Gambar 4. 10 Grafik Waktu Penipuan Transaksi	50
Gambar 4. 11 Total Transaksi Unik.....	51
Gambar 4. 12 Dataset setelah penghapusan kolom tidak terpakai.....	52
Gambar 4. 13 Dataset Sebelum Pembersihan Column Merchant.....	53
Gambar 4. 14 Dataset Setelah Pembersihan Column Merchant.....	53
Gambar 4. 15 Output Head Setelah Encoding	54
Gambar 4. 16 Distribusi Data Class sebelum dilakukan balancing data	55
Gambar 4. 17 Distribusi Data Class setelah dilakukan balancing data.....	55
Gambar 4. 18 Split data train-test	56
Gambar 4. 19 Parameter terbaik untuk Hyperparameter Tuning.....	58
Gambar 4. 20 Output Hasil Akurasi Seluruh Model.....	60
Gambar 4. 21 Confution Matrix Logistic Regression.....	63
Gambar 4. 22 Confution Matrix Decision Tree	63
Gambar 4. 23 Confution Matrix Random Forest	63
Gambar 4. 24 Confution Matrix Naïve Bayes	64
Gambar 4. 25 Confution Matrix XG Boost	64
Gambar 4. 26 Confution Matrix SVM.....	64

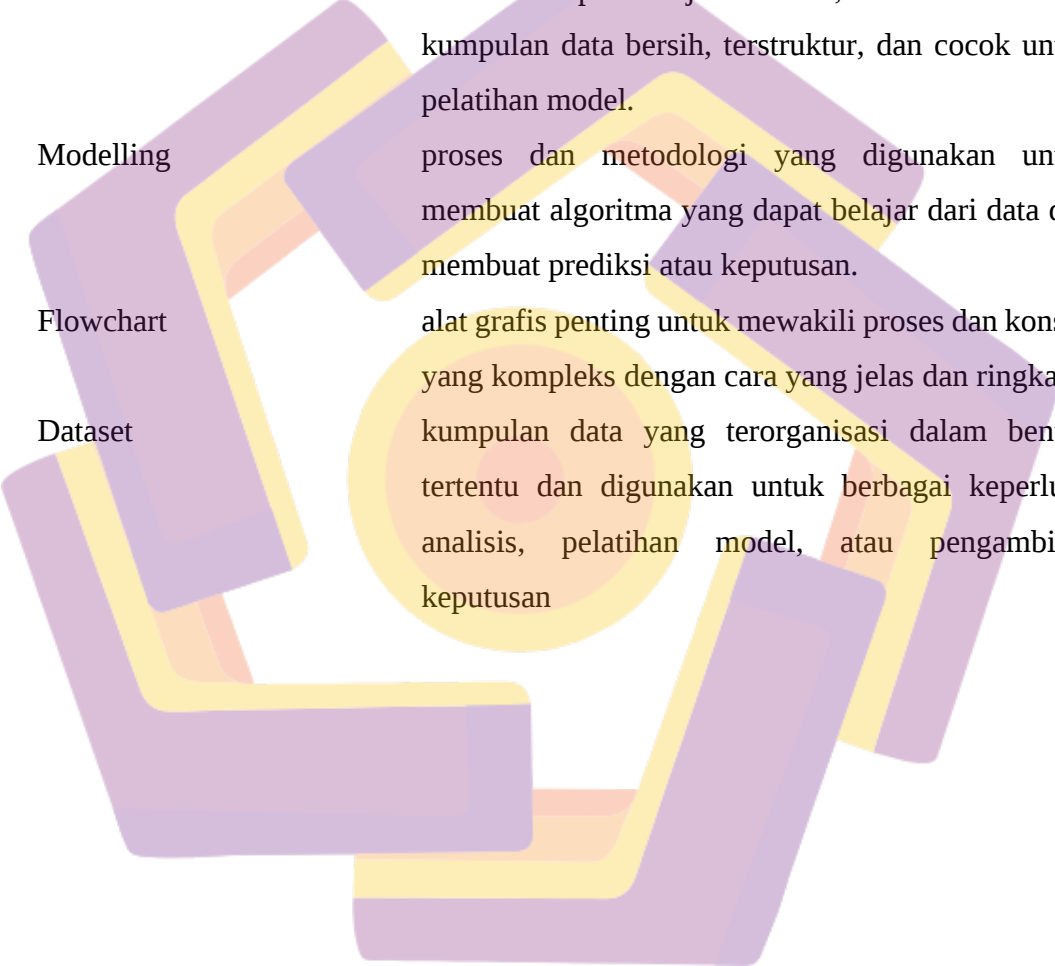
DAFTAR LAMPIRAN

Lampiran 1 Tampilan website kaggle	75
Lampiran 2 Credit Card Transactions Fraud Detection Dataset	75
Lampiran 3 Hasil Model Logistic Regression Sebelum Hyperparameter Tuning	76
Lampiran 4 Hasil Model Logistic Regression Sesudah Hyperparameter Tuning	76
Lampiran 5 Hasil Evaluasi Model Decision Tree Sebelum Hyperparameter Tuning	76
Lampiran 6 . Hasil Evaluasi Model Random Forest Sebelum Hyperparameter Tuning	77
Lampiran 7 Hasil Model Random Forest Sesudah Hyperparameter Tuning	93
Lampiran 8 Hasil Evaluasi Model Naïve Bayes Sebelum Hyperparameter Tuning	77
Lampiran 9 Hasil Evaluasi Model Naïve Bayes Sesudah Hyperparameter Tuning	77
Lampiran 10 Hasil Evaluasi Model XGBoost Sebelum Hyperparameter Tuning	78
Lampiran 11 Hasil Evaluasi Model XGBoost Sesudah Hyperparameter Tuning	78
Lampiran 12 Hasil Evaluasi Model SVM Sebelum Hyperparameter Tuning	78
Lampiran 13 Hasil Evaluasi Model SVM Sesudah Hyperparameter Tuning	78

DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
RF	Random Forest
XGBoost	Extreme Gradient Boost
LR	Logistic Regression
DT	Decision Tree
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
EDA	Exploratory Data Analysis
AUC	Area Under the Curve

DAFTAR ISTILAH



Algoritma	urutan instruksi terstruktur yang dirancang untuk memecahkan masalah tertentu secara efisien dan akurat.
Preprocessing	langkah penting dalam mempersiapkan data untuk analisis dan pembelajaran mesin, memastikan bahwa kumpulan data bersih, terstruktur, dan cocok untuk pelatihan model.
Modelling	proses dan metodologi yang digunakan untuk membuat algoritma yang dapat belajar dari data dan membuat prediksi atau keputusan.
Flowchart	alat grafis penting untuk mewakili proses dan konsep yang kompleks dengan cara yang jelas dan ringkas
Dataset	kumpulan data yang terorganisasi dalam bentuk tertentu dan digunakan untuk berbagai keperluan analisis, pelatihan model, atau pengambilan keputusan

INTISARI

Fraud atau kecurangan dalam transaksi e-commerce merupakan tantangan besar yang dapat menyebabkan kerugian finansial bagi perusahaan dan konsumen. Beberapa studi sebelumnya telah mengembangkan metode deteksi fraud menggunakan algoritma Machine Learning, namun masih terdapat kendala seperti ketidakseimbangan data serta akurasi model yang belum optimal. Oleh karena itu, penelitian ini menerapkan beberapa algoritma Machine Learning, yaitu Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, XGBoost, dan Support Vector Machine (SVM) untuk mendeteksi transaksi fraud berdasarkan dataset yang digunakan.

Metode penelitian yang digunakan dalam penelitian ini meliputi preprocessing data, balancing data, serta pelatihan dan evaluasi model Machine Learning. Dataset yang digunakan diproses dengan teknik data cleaning, encoding data kategori, normalisasi, dan penyeimbangan data menggunakan teknik SMOTE dan SMOTETomek guna mengatasi ketidakseimbangan data antara transaksi fraud dan non-fraud. Model Machine Learning diuji menggunakan teknik cross-validation dan dievaluasi berdasarkan metrik akurasi, presisi, recall, f1-score, dan AUC untuk menentukan performa terbaik dalam mendeteksi transaksi fraud.

Hasil penelitian menunjukkan bahwa XGBoost menjadi algoritma dengan performa terbaik dengan nilai akurasi sebesar 0.99, recall sebesar 0.82, presisi sebesar 0.93, f1-score sebesar 0.87, dan AUC sebesar 0.99. Algoritma Decision Tree juga menunjukkan hasil yang cukup baik dengan akurasi 0.99 dan f1-score 0.81. Sementara itu, Random Forest memiliki akurasi 0.98 namun dengan presisi yang lebih rendah, yaitu 0.20. Penelitian ini berkontribusi dalam pengembangan sistem deteksi fraud yang lebih akurat untuk meningkatkan keamanan transaksi e-commerce.

Kata kunci: Deteksi Fraud, Machine Learning, XGBoost, Random Forest, SMOTE, SMOTETomek.

ABSTRACT

Fraud in e-commerce transactions is a major challenge that can lead to financial losses for both companies and consumers. Several previous studies have developed fraud detection methods using Machine Learning algorithms; however, there are still challenges such as data imbalance and suboptimal model accuracy. Therefore, this study applies several Machine Learning algorithms, namely Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, XGBoost, and Support Vector Machine (SVM), to detect fraudulent transactions based on the dataset used.

The research methodology includes data preprocessing, data balancing, and training and evaluation of Machine Learning models. The dataset is processed using data cleaning, categorical data encoding, normalization, and data balancing techniques using SMOTE and SMOTETomek to address the imbalance between fraudulent and non-fraudulent transactions. The Machine Learning models are tested using cross-validation techniques and evaluated based on accuracy, precision, recall, f1-score, and AUC metrics to determine the best performance in detecting fraud.

The results show that XGBoost is the best-performing algorithm, achieving an accuracy of 0.99, recall of 0.82, precision of 0.93, f1-score of 0.87, and AUC of 0.99. The Decision Tree algorithm also performs well with an accuracy of 0.99 and f1-score of 0.81. Meanwhile, Random Forest achieves an accuracy of 0.98 but with a lower precision of 0.20. This study contributes to the development of more accurate fraud detection systems to enhance the security of e-commerce transactions.

Keyword: Fraud Detection, Machine Learning, XGBoost, Random Forest, SMOTE, SMOTETomek.