

**PENERAPAN ALGORITMA REGRESI LOGISTIK DAN RANDOM
FOREST DALAM MENDETEKSI TRANSAKSI FRAUD**

JALUR PROFESIONAL – STUDI INDEPENDEN

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



disusun oleh

ANANG SANDRHYNO BADRAHAM

19.12.1354

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

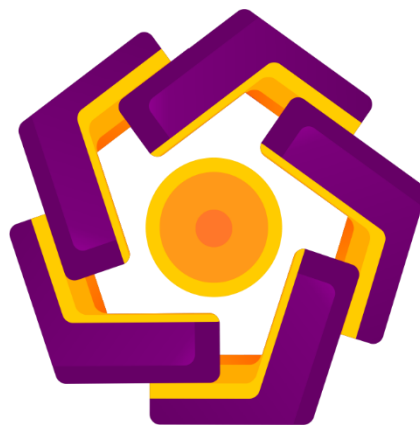
2022

**PENERAPAN ALGORITMA REGRESI LOGISTIK DAN RANDOM
FOREST DALAM MENDETEKSI TRANSAKSI FRAUD**

JALUR PROFESIONAL – STUDI INDEPENDEN

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1 Sistem Informasi



disusun oleh

ANANG SANDRHYNO BADRAHAM

19.12.1354

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

HALAMAN PERSETUJUAN
JALUR PROFESIONAL – STUDI INDEPENDEN
PENERAPAN ALGORITMA REGRESI LOGISTIK DAN RANDOM FOREST
DALAM MENDETEKSI TRANSAKSI FRAUD

yang disusun dan diajukan oleh

Anang Sandrhyno Badraham

19.12.1354

telah disetujui oleh Dosen Pembimbing
pada tanggal 3 November 2022

Dosen Pembimbing,



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

HALAMAN PENGESAHAN
JALUR PROFESIONAL – STUDI INDEPENDEN
PENERAPAN ALGORITMA REGRESI LOGISTIK DAN RANDOM FOREST
DALAM MENDETEKSI TRANSAKSI FRAUD

yang disusun dan diajukan oleh

Anang Sandrhyno Badraham

19.12.1354

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 November 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Bety Wulan Sari, M.Kom

NIK. 190302254

Subektiningsih, M.Kom

NIK. 190302413

Yoga Pristvanto, S.Kom, M.Eng

NIK. 190302412

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 November 2022

DEKAN FAKULTAS ILMU KOMPUTER

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

NIK. 190302096

HALAMAN PERNYATAAN KONTRIBUSI KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Anang Sandrhyno Badraham

NIM : 19.12.1354

Menyatakan bahwa karya dengan judul berikut:

Penerapan Algoritma Regresi Logistik dan Random Forest Dalam Mendeteksi Transaksi Fraud

Dosen Pembimbing : Yoga Pristyanto, S.Kom., M.Eng

1. Karya adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya merupakan gagasan penelitian yang orisinal dan **SAYA** memiliki **KONTRIBUSI** terhadap karya tersebut.
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 atau Referensi 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, 4 November 2022

Yang Menyatakan,



Anang Sandrhyno Badraham

HALAMAN PERSEMBAHAN

Puji syukur peneliti panjatkan kehadiran Allah SWT yang telah memberikan karunia dan limpahan kasih dan rahmat-Nya sehingga peneliti dapat menyelesaikan skripsi dengan judul “Penerapan Algoritma Regresi Logistik dan Random Forest Dalam Mendeteksi Transaksi Fraud. Peneliti menyadari bahwa tersusun dan terselesaikannya skripsi ini tidak lepas dari banyak pihak yang memberikan bantuan. Maka dari itu peneliti mengucapkan terima kasih kepada:

1. Kedua orang tua yaitu Abah Anang Sandi Risnopandi dan Mamah Baiq Yunita Haryanti yang tiada henti-hentinya memberikan doa, dukungan, saran, perhatian, kesabaran, keikhlasan serta bantuan dalam segi fisik dan material selama proses menyelesaikan perkuliahan jenjang S1 Sistem Infomasi ini.
2. Kedua Adikku, Galuh Inge Xhavaira dan Anang Rayyan Dessandyo Badraham yang selalu memberikan motivasi, doa, dan semangat yang luar biasa untuk selalu belajar sehingga skripsi ini dapat diselesaikan.
3. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta
4. Bapak Anggit Dwi Hartanto, M.Kom. selaku Kepala Program Studi Sistem Informasi Universitas Amikom Yogyakarta.
5. Bapak Yoga Pristyanto, S.Kom., M.Eng. selaku Dosen Pembimbing Skripsi Program Studi Sistem Informasi Universitas Amikom Yogyakarta yang selalu memberikan dan meluangkan waktunya untuk membimbing dan mengarahkan dengan penuh kesabaran.

Semoga semua arahan dan bimbingan serta doa dari semua pihak diberikan balasan oleh Allah SWT segala amal baik yang telah diberikan. Atas perhatiannya peneliti mengucapkan terimakasih.

KATA PENGANTAR

Penulis bersyukur kepada Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "Penerapan Algoritma Regresi Logistik dan Random Forest Dalam Mendeteksi Transaksi Fraud", dalam rangka memenuhi syarat untuk memperoleh gelar sarjana Program Studi Sistem Informasi. Dalam pembuatan skripsi ini tidak lepas dari bantuan pihak-pihak, untuk itu dalam kesempatan ini penulis mengucapkan terima kasih kepada:

1. Bapak Anggit Dwi Hartanto, M.Kom. selaku Kepala Program Studi S1 Sistem Informasi Universitas Amikom Yogyakarta,
2. Bapak Yoga Pristyanto, S.Kom., M.Eng. selaku Dosen Pembimbing Skripsi Program Studi Sistem Informasi Universitas Amikom Yogyakarta.
3. Almira Syahadati Arsyah, Ayu Sulistiara, Iraningtyas Parasetiyani, dan Tegar Abdillah selaku teman kelompok proyek akhir Studi Independen.
4. Kedua orangtua dan teman-teman yang mendukung dan memberikan semangat serta nasihat dalam menyelesaikan tulisan ini, serta semua pihak yang tidak dapat disebutkan satu persatu.

Penulis sadar sepenuhnya bahwa skripsi ini masih jauh dari sempurna, untuk itu kritik dan saran membangun sangat diperlukan untuk menyempurnakan skripsi ini.

Yogyakarta, 4 November 2022

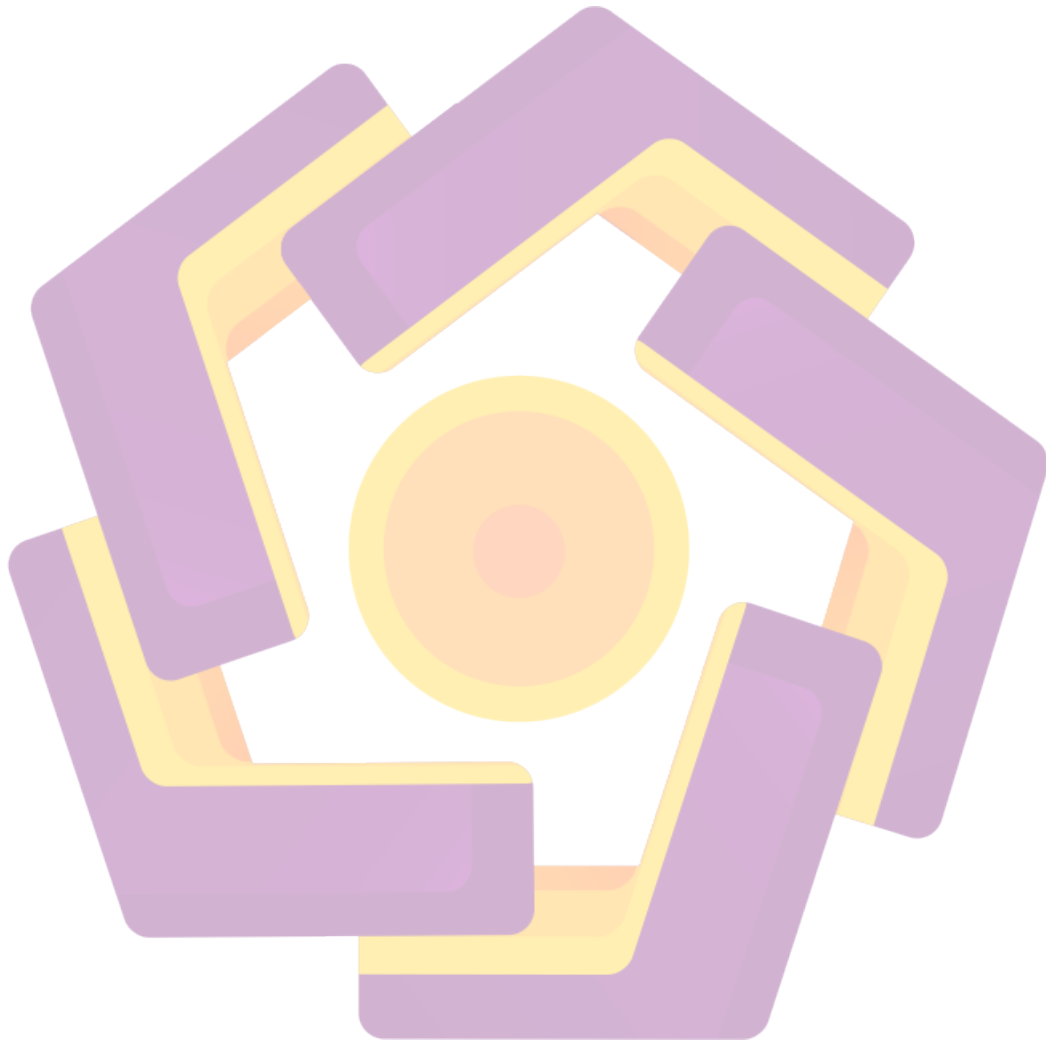
Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KONTRIBUSI KARYA.....	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	viii
DAFTAR TABEL.....	viii
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
INTISARI	xii
ABSTRACT	xiii
BAB I PENDAHULUAN.....	14
1.1 Latar Belakang	14
1.2 Batasan Masalah	14
1.3 Tujuan Penelitian	15
1.4 Manfaat Penelitian	15
1.5 Profil	15
1.6 Landasan Teori	17
BAB II PEMBAHASAN	20
2.1 Alur Pengembangan Produk	20
2.2 Analisis Masalah Teknis dan Penyelesaian	21
2.3 Pembahasan Produk	21
2.4 Hasil Proyek Akhir	32
2.5 Pembahasan Kegiatan	38
2.6 Peran dan Kontribusi	40
BAB III PENUTUP	41
3.1 Kesimpulan	41
3.2 Saran	41
REFERENSI	42
LAMPIRAN	43

DAFTAR TABEL

Tabel 2. 1 Masalah Teknis dan Penyelesaian	21
Tabel 2. 2 Jadwal Pembelajaran dan Topik Pembelajaran	38
Tabel 2. 3 Peran dan Kontribusi	40

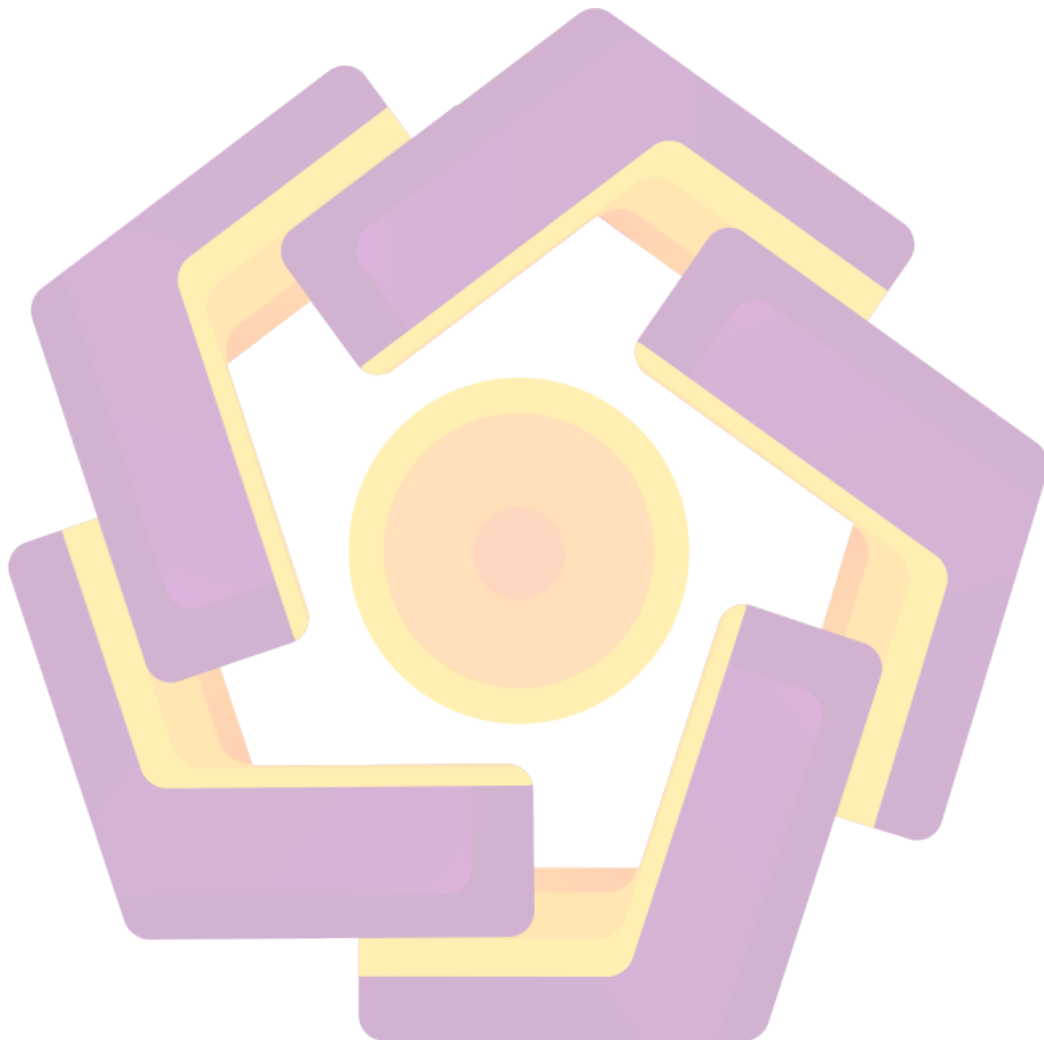


DAFTAR GAMBAR

Gambar 2. 1 Alur Pengembangan	20
Gambar 2. 2 Import library	22
Gambar 2. 3 Lima dataset teratas	22
Gambar 2. 4 Jumlah row dan column dataset	22
Gambar 2. 5 Memeriksa tipe data	23
Gambar 2. 6 Mengubah tipe data	23
Gambar 2. 7 Tipe data setelah diubah	24
Gambar 2. 8 Memeriksa duplikasi data	24
Gambar 2. 9 Memeriksa kesamaan antara kolom 1	25
Gambar 2. 10 Drop kolom 'customerId'	25
Gambar 2. 11 Memeriksa kesamaan antara kolom 2	25
Gambar 2. 12 Mengisi blank space dengan NaN	26
Gambar 2. 13 Memeriksa jumlah missing value	26
Gambar 2. 14 Memeriksa kolom dengan missing value lebih dari 70%	27
Gambar 2. 15 Menghapus kolom dengan missing value lebih dari 70%	27
Gambar 2. 16 Mengisi kolom dengan missing value	27
Gambar 2. 17 Histogram data numerikal	27
Gambar 2. 18 Boxplot kolom 'creditLimit' sebelum outliers handling	28
Gambar 2. 19 Boxplot kolom 'creditLimit' setelah outliers handling	28
Gambar 2. 20 Delapan variabel yang memiliki nilai korelasi tinggi	28
Gambar 2. 21 Model Regresi logistik	29
Gambar 2. 22 Model Regresi Logistik + SMOTE	30
Gambar 2. 23 Model Random Forest	31
Gambar 2. 24 Model Random Forest + SMOTE	31
Gambar 2. 25 hasil analisa fraud menggunakan regresi logistik	32
Gambar 2. 26 hasil analisa fraud menggunakan regresi logistik + SMOTE	33
Gambar 2. 27 hasil analisa fraud menggunakan Random Forest	34
Gambar 2. 28 hasil analisa fraud menggunakan Random Forest + SMOTE	35
Gambar 2. 29 Dashboard Visualisasi	36
Gambar 2. 30 Credit Card Fraud Detection App	37

DAFTAR LAMPIRAN

Lampiran 1. <i>Logbook</i>	43
Lampiran 2. Sertifikat dan Transkrip Nilai Studi Independen.....	49
Lampiran 3. Kartu Rencana Studi.....	50



INTISARI

Fraud adalah penipuan yang disengaja untuk mendapat keuntungan. Upaya untuk mendeteksi penipuan transaksi keuangan dapat menggunakan *machine learning* dengan mengambil data transaksi keuangan secara elektronik. Hal ini dapat membantu mencegah terjadinya penipuan. Tujuan penelitian ini diantaranya untuk mengetahui *feature important* atau variabel apa saja yang berpengaruh signifikan terhadap variabel dependen yaitu *fraud*, untuk mengetahui nilai akurasi klasifikasi label *fraud* dengan algoritma regresi logistik dan *random forest* dan untuk memprediksi transaksi yang telah terjadi dengan aplikasi berbasis website. Untuk melakukan deteksi penipuan digunakan algoritma regresi logistik dan *random forest* melalui lima langkah: *business understanding, data preparation, visualisasi, modeling dan deploying website*. Variabel yang berpengaruh signifikan terhadap variabel *fraud* adalah *Transaction Amount, Pos Entry Mode, Card Present, MatchCVV, Transaction Type*. Pemodelan regresi logistik memiliki performa yang baik dengan nilai akurasi 98%, pemodelan regresi logistik dengan menggunakan *SMOTE* mendapat nilai akurasi sebesar 73%. Sementara pemodelan *random forest* juga memiliki performa yang baik dengan nilai akurasi 97% dan pemodelan *random Forest* menggunakan *SMOTE* mendapatkan nilai akurasi sebesar 55% dan Aplikasi berbasis website dengan nama *Credit Card Fraud Detection* dapat memprediksi transaksi *fraud* setelah terjadinya sebuah transaksi.

Kata kunci: *Fraud, Machine learning, Regresi logistik, Random Forest*

ABSTRACT

Fraud is a deliberate deception for profit. Efforts to detect fraudulent financial transactions can use machine learning by retrieving financial transaction data electronically. This can help prevent fraud from occurring. The purposes of this study include to find out the important features or variables that have a significant effect on the dependent variable, namely fraud, to find out the accuracy of the fraud label classification using logistic regression algorithms and random forest and to predict transactions that have occurred with website-based applications. To detect fraud, logistic regression algorithms and random forest through five steps: business understanding, data preparation, visualization, and modeling. Variables that have a significant effect on the fraud variable are Transaction Amount, Post Entry Mode, Card Present, MatchCVV, Transaction Type. Logistic regression modeling has a good performance with an accuracy value of 98% and logistic regression modeling using SMOTE has an accuracy value of 73%. Meanwhile random forest also has good performance with an accuracy value of 97% and random forest modeling using SMOTE gets an accuracy value of 55% and a website-based application called Credit Card Fraud Detection can predict fraud transactions after a transaction has occurred.

Keyword: Fraud, Machine learning, Logistic Regression, Random Forest