

TESIS

**DETEKSI *FRAUD* PADA TRANSAKSI KEUANGAN
MENGUNAKAN METODE *STACKING ENSEMBLE
LEARNING***



WILDAN KRISTIAN MAHARDIKA
No. Mhs.: 225312020/PS/MTF

PROGRAM STUDI MAGISTER INFORMATIKA
PROGRAM PASCASARJANA
UNIVERSITAS ATMA JAYA YOGYAKARTA
2023

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

DETEKSI FRAUD PADA TRANSAKSI KEUANGAN MENGGUNAKAN METODE STACKING ENSEMBLE
LEARNING

yang disusun oleh

Wildan Kristian Mahardika

225312020

dinyatakan telah memenuhi syarat pada tanggal 23 Oktober 2023

		Keterangan
Dosen Pembimbing 1	: Paulus Mudjihartono, S.T., M.T., Ph.D	Telah Menyetujui
Dosen Pembimbing 2	: Dr. Pranowo, S.T., M.T.	Telah Menyetujui
Tim Penguji		
Penguji 1	: Paulus Mudjihartono, S.T., M.T., Ph.D	Telah Menyetujui
Penguji 2	: Dra. Ernawati, M.T., Ph.D	Telah Menyetujui

Yogyakarta, 23 Oktober 2023

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Dr. Ir. Parama Kartika Dewa SP., S.T., M.T.

Dokumen ini merupakan dokumen resmi UAJY yang tidak memerlukan tanda tangan karena dihasilkan secara elektronik oleh Sistem Bimbingan UAJY. UAJY bertanggung jawab penuh atas informasi yang tertera di dalam dokumen ini

HALAMAN PERNYATAAN TESIS

Saya yang bertandatangan dibawah ini:

Nama : WILDAN KRISTIAN MAHARDIKA

Nomor Mahasiswa : 225312020/PS/MTF

Judul Tesis : DETEKSI FRAUD PADA TRANSAKSI KEUANGAN
MENGUNAKAN METODE STACKING ENSEMBLE
LEARNING

Menyatakan dengan sesungguhnya bahwa Tesis ini merupakan hasil penelitian, pemikiran dan pemaparan asli saya sendiri. Saya tidak mencantumkan tanpa pengakuan bahan - bahan yang telah dipublikasikan sebelumnya atau ditulis oleh orang lain, atau sebagai bahan yang pernah diajukan untuk gelar atau ijasah pada Universitas Atma Jaya Yogyakarta atau perguruan tinggi lainnya.

Apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik sesuai dengan peraturan yang berlaku di Universitas Atma Jaya Yogyakarta.

Demikian pernyataan ini saya buat.

Yogyakarta, 09 Oktober 2023

Yang membuat pernyataan,

(WILDAN KRISTIAN MAHARDIKA)

INTISARI

Financial services merupakan sebuah sektor yang bergerak di bidang layanan keuangan dengan tujuan memudahkan masyarakat mengelola aset keuangannya. Dalam kurun waktu tahun 2017 hingga 2021, tercatat bahwa kasus kejahatan di sektor keuangan mengalami kenaikan dari 370.916 kasus menjadi 1.686.121 kasus. Salah satu penyebab terjadinya kejahatan tersebut adalah pada penipuan transaksi pelanggan. Cara yang dapat dilakukan untuk mengatasi masalah tersebut adalah dengan melakukan deteksi di setiap transaksi. Deteksi transaksi yang dilakukan secara manual dan konvensional rawan terjadi *human error*. Selain itu, semakin berkembangnya data juga membuat deteksi secara manual menjadi tidak efektif.

Penelitian ini mengusulkan sebuah metode untuk mendeteksi transaksi di sebuah *financial services*. Deteksi *fraud* dilakukan menggunakan teknik *stacking ensemble learning*. Peneliti akan mengkombinasikan tiga algoritma *machine learning*, yaitu *Support Vector Machine (SVM)*, *Logistic Regression (LR)*, dan *k-Nearest Neighbour (kNN)*. Sedangkan algoritma *stacking ensemble learning* yang digunakan yaitu *Random Forest*. Untuk kasus *fraud detection*, terjadi ketidakseimbangan kelas antara kelas *fraud* dan kelas *non-fraud*. Metode SMOTE digunakan untuk mengatasi masalah ketidakseimbangan kelas pada dataset.

Hasil pengujian dilakukan menggunakan *confusion matrix*. Hasil membuktikan bahwa dengan menggunakan *stacking ensemble learning* dapat meningkatkan akurasi deteksi model dasar, yaitu sejumlah 0,14% untuk LR, 0,47% untuk SVM, dan 0,13% untuk kNN. Hasil performa *stacking ensemble learning* memiliki akurasi sebesar 99,78%, presisi sebesar 99,81%, *recall* sebesar 99,96%, dan f1-score sebesar 99,88%.

Kata kunci: Deteksi Fraud, SMOTE, Ensemble Learning, Imbalanced Dataset, Stacking

ABSTRACT

Financial services is a sector with the aim of making it easier for people to manage their financial assets. In the period 2017 to 2021, it was recorded that crime cases in the financial sector increased from 370,916 cases to 1,686,121 cases. One of the causes of this crime is fraudulent customer transactions. One way to overcome this problem is to detect every transaction. Detection of transactions carried out manually and conventionally is prone to human error. Apart from that, the increasing development of data also makes manual detection ineffective.

This research proposes a method for detecting transactions in financial services. Fraud detection is carried out using stacking ensemble learning techniques. Researchers will combine three machine learning algorithms, namely Support Vector Machine (SVM), Logistic Regression (LR), and k-Nearest Neighbour (kNN). Meanwhile, the stacking ensemble learning algorithm used is Random Forest. For fraud detection cases, there is a class imbalance between the fraud class and the non-fraud class. The SMOTE method is used to overcome the problem of class imbalance in the dataset.

Test results were carried out using a confusion matrix. The results prove that using stacking ensemble learning can increase the detection accuracy of the basic model, namely 0.14% for LR, 0.47% for SVM, and 0.13% for kNN. The stacking ensemble learning performance results have an accuracy of 99.78%, precision of 99.81%, recall of 99.96%, and f1-score of 99.88%.

Keywords: Fraud Detection, SMOTE, Ensemble Learning, Imbalanced Dataset, Stacking

KATA PENGANTAR

Segala puji dan syukur kepada Tuhan Yesus yang telah menyertai penulis selama melakukan penelitian dan menyusun laporan tesis ini sehingga penulisan laporan tesis ini dapat selesai dengan baik dan lancar. Penulisan tesis ini menjadi syarat wajib bagi mahasiswa pascasarjana Universitas Atma Jaya Yogyakarta untuk memperoleh gelar Magister Komputer (M.Kom.) pada program studi Magister Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta. Penulis berharap supaya melalui penelitian ini dapat memberi manfaat bagi semua pihak serta dapat dikembangkan lebih lanjut lagi.

Selama melakukan penelitian ini, penulis telah menerima banyak bantuan dan bimbingan dari banyak pihak, baik secara langsung maupun tidak langsung. Dengan kerendahan hati, penulis mengucapkan terimakasih kepada:

1. Tuhan Yesus Kristus. Atas penyertaanNya sehingga penelitian ini dapat selesai dengan baik dan lancar.
2. Kedua orangtua yang selalu mendoakan, memberi semangat, motivasi, dukungan kepada penulis selama pengerjaan tesis ini.

Serta kepada kakak saya, mas Bernadiaz Danuarta, S.E. yang terus memberi semangat dan dukungan kepada penulis.

3. Bp. Paulus Mudjihartono, S.T., M.T., Ph.D selaku dosen pembimbing pertama dan Bp. Dr. Pranowo, S.T., M.T. selaku dosen pembimbing kedua yang telah membimbing, memberi solusi, memberi motivasi selama pengerjaan tesis ini.

4. Seluruh dosen prodi Magister Informatika yang telah memberikan ilmu yang sangat berguna kepada penulis. Juga yang telah mendidik, mengajar penulis selama kuliah.
5. Teman-teman Magister Informatika angkatan 2022 yang telah memberi semangat dan dukungan selama kuliah hingga menyusun penelitian tesis ini.
6. Kepada teman dekat penulis, yaitu Aryasasta Dharma Sagala, Riko Gusjakal Muran, Kadek Doni Hendrawan, Wegig Arok Priambodo, Nathania Saphira yang selalu mendukung dan memberi ide-ide kepada penulis.
7. Semua pihak yang tidak dapat disebutkan satu persatu, yang telah memberikan dukungan dan semangat kepada penulis, baik secara langsung maupun tidak langsung.

Akhir kata, penulis mengucapkan terimakasih kepada seluruh pihak yang telah membantu dalam penelitian dan penyusunan laporan tesis ini serta memohon maaf apabila masih terlalu banyak kekurangan yang terjadi. Semoga dengan adanya penelitian ini, dapat memberikan manfaat bagi semua pihak dan tentunya bagi kemajuan teknologi informasi di Indonesia.

Yogyakarta, 09 Oktober 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN TESIS.....	iv
INTISARI.....	v
ABSTRACT.....	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Keaslian Penelitian	5
1.5 Manfaat Penelitian.....	5
1.6 Tujuan Penelitian.....	6
1.7 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA.....	8
BAB III LANDASAN TEORI.....	21
3.1 Transaksi pada Layanan Keuangan	21

3.2	<i>Fraud Detection</i>	22
3.3	<i>Machine Learning</i>	23
3.3.1	<i>K-Nearest-Neighbour (KNN)</i>	25
3.3.2	<i>Support Vector Machine (SVM)</i>	26
3.3.3	<i>Logistic Regression (LR)</i>	27
3.3.4	<i>Random Forest (RF)</i>	28
3.4	<i>Ensemble Learning</i>	30
3.4.1	<i>Stacking Ensemble Learning</i>	30
3.5	<i>Imbalanced Dataset</i>	32
3.5.1	<i>Synthetic Minority Oversampling Technique (SMOTE)</i>	33
3.6	<i>Performance Evaluation</i>	34
3.6.1	<i>Confusion Matrix</i>	34
BAB IV METODOLOGI PENELITIAN		37
4.1	Tahapan Penelitian	37
4.2	Metode Penelitian	39
4.2.1	Studi Literatur	39
4.2.2	<i>Data collection</i>	40
4.2.3	<i>Imbalanced dataset processing</i>	41
4.2.4	<i>Splitting dataset</i>	42
4.2.5	Pembuatan model algoritma	43

4.2.6	<i>Testing model</i>	43
BAB V HASIL DAN PEMBAHASAN		45
5.1	<i>Imbalanced Dataset Processing</i>	45
5.2	Pembuatan Model Algoritma	47
5.2.1	<i>Base Model</i>	48
5.2.2	<i>Meta Model</i>	52
5.3	Testing Model	54
5.3.1	<i>Base Model Logistic Regression</i>	54
5.3.2	<i>Base Model Support Vector Machine</i>	55
5.3.3	<i>Base Model k-Nearest-Neighbour</i>	56
5.3.4	<i>Meta Model Random Forest</i>	57
5.4	Implementasi Model pada <i>Real Case</i>	60
BAB VI KESIMPULAN DAN SARAN		64
6.1	Kesimpulan	64
6.2	Saran	65
DAFTAR PUSTAKA		66

DAFTAR TABEL

Tabel 1. Daftar Penelitian Terdahulu	18
Tabel 2. Nama kolom dan deskripsi dataset.....	40
Tabel 3. Perbandingan jumlah sampling menggunakan SMOTE.....	46
Tabel 4. Rekap hasil pengujian model	59

DAFTAR GAMBAR

Gambar 1. Fungsi Sigmoid	27
Gambar 2. Tahapan <i>Random Forest</i>	29
Gambar 3. Tabel Confusion Matrix	35
Gambar 4. Flowchart Penelitian.....	38
Gambar 5. Sample dataset.....	41
Gambar 6. Model Stacking Ensemble Learning	43
Gambar 7. Perbandingan data fraud dan data non-fraud.....	45
Gambar 8. Parameter SMOTE up-sampling	46
Gambar 9. Parameter operator split validation	48
Gambar 10. Proses awal pembuatan model dengan kondisi imbalanced dataset .	49
Gambar 11. Proses awal pembuatan model menggunakan metode SMOTE.....	49
Gambar 12. Proses pembuatan model dasar logistic regression	50
Gambar 13. Proses pembuatan model dasar support vector machine	50
Gambar 14. Proses konversi tipe data pada SVM.....	51
Gambar 15. Proses pembuatan model dasar k-NN	51
Gambar 16. Tahap split dataset pada meta model.....	52
Gambar 17. Proses pembuatan model stacking ensemble learning	52
Gambar 18. Parameter algoritma Random Forest.....	53
Gambar 19. Confusion Matrix Logistic Regression Imbalanced Dataset.....	54
Gambar 20. Confusion Matrix Logistic Regression menggunakan SMOTE 50:50	55
Gambar 21. Confusion Matrix Support Vector Machine Imbalanced Dataset	55

Gambar 22. Confusion Matrix Support Vector Machine menggunakan SMOTE 50:50.....	56
Gambar 23. Confusion Matrix k-Nearest-Neighbour Imbalanced Dataset.....	56
Gambar 24. Confusion Matrix k-Nearest-Neighbour menggunakan SMOTE 50:50	57
Gambar 25. Confusion Matrix Meta Model Random Forest Imbalanced Dataset	58
Gambar 26. Confusion Matrix meta model Random Forest menggunakan SMOTE 50:50.....	58
Gambar 27. Proses pelatihan model (berlabel)	61
Gambar 28. Proses prediksi model (tidak berlabel).....	62
Gambar 29. Hasil prediksi untuk dataset yang tidak berlabel.....	62