

**MODEL HIBRIDA SVM-RANDOM FOREST UNTUK
OPTIMASI KLASIFIKASI SENTIMEN PENGADUAN PUBLIK
PADA FITUR LAKSA APLIKASI TANGERANG LIVE**



SKRIPSI

Maura Davita Erzanda

00000088620

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**MODEL HIBRIDA SVM-RANDOM FOREST UNTUK
OPTIMASI KLASIFIKASI SENTIMEN PENGADUAN PUBLIK
PADA FITUR LAKSA APLIKASI TANGERANG LIVE**



SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer

Maura Davita Erzanda

00000088620

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Maura Davita Erzanda

Nomor Induk Mahasiswa : 00000088620

Program Studi : Sistem Informasi

Skripsi dengan judul:

**MODEL HIBRIDA SVM-RANDOM FOREST UNTUK
OPTIMASI KLASIFIKASI SENTIMEN PENGADUAN PUBLIK
PADA FITUR LAKSA APLIKASI TANGERANG LIVE**

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 28 November 2025



Maura Davita Erzanda

LEMBAR PERNYATAAN PENGGUNAAN BANTUAN

KECERDASAN ARTIFISIAL (AI)

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Maura Davita Erzanda
NIM : 00000088620
Program Studi : Sistem Informasi
Judul Skripsi : Model Hibrida SVM-Random Forest untuk Optimasi Klasifikasi Sentimen Pengaduan Publik pada Fitur Laksa Aplikasi Tangerang LIVE

Dengan ini saya menyatakan secara jujur menggunakan bantuan Kecerdasan Artifisial (AI) dalam pengerjaan Tugas/Project/Tugas Akhir*(coret salah satu) sebagai berikut (beri tanda centang yang sesuai):

- ✓ Menggunakan AI sebagaimana diizinkan untuk membantu dalam menghasilkan ide-ide utama saja
- ✓ Menggunakan AI sebagaimana diizinkan untuk membantu menghasilkan teks pertama saja
- ✓ Menggunakan AI untuk menyempurnakan sintaksis dan tata bahasa untuk pengumpulan tugas
- ☐ Karena tidak diizinkan: Tidak menggunakan bantuan AI dengan cara apa pun dalam pembuatan tugas

Saya juga menyatakan bahwa:

- (1) Menyerahkan secara lengkap dan jujur penggunaan perangkat AI yang diperlukan dalam tugas melalui Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)
- (2) Saya mengakui bahwa saya telah menggunakan bantuan AI dalam tugas saya baik dalam bentuk kata, paraphrase, penyertaan ide atau fakta penting yang disarankan oleh AI dan saya telah menyantumkan dalam sitasi serta referensi

Terlepas dari pernyataan di atas, tugas ini sepenuhnya merupakan karya saya sendiri

Tangerang, 28 November 2025
Yang Menyatakan,



Maura Davita Erzanda

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Maura Davita Erzanda

NIM : 00000088620

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah : Perbandingan Algoritma Support Vector Machine (SVM) dan Random Forest Untuk Analisis Sentimen Pengaduan Publik Dalam Fitur Laksa Aplikasi Tangerang Live

Menyatakan dengan sesungguhnya bahwa saya bersedia* (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 28 November 2025



(Maura Davita Erzanda)

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi ini dengan judul “Model Hibrida SVM-Random Forest untuk Optimasi Klasifikasi Ssentimen Pengaduan Publik pada Fitur LAKSA Aplikasi Tangerang LIVE”.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer S.Kom pada Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara.

Penulis menyadari bahwa dalam proses penyusunan skripsi ini, penulis banyak mendapatkan bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Ir. Andrey Andoko, M.Sc., Ph.D selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc. selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom. selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara sekaligus pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesaikannya tugas akhir ini.
4. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
5. Teman-teman saya yang senantiasa setia menemani, mendengarkan, dan memberikan dukungan selama proses penulisan tugas akhir ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, baik dari segi materi maupun tata bahasa. Oleh karena itu, penulis sangat mengharapkan

kritik dan saran yang membangun dari semua pihak demi penyempurnaan skripsi ini di masa mendatang.

Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilm pengetahuan, khususnya di bidang *Data Science* dan *Machine learning* serta dapat menjadi referensi bagi penelitian selanjutnya.

Tangerang, 28 November 2025



(Maura Davita Erzanda)



MODEL HIBRIDA SVM-RANDOM FOREST UNTUK OPTIMASI KLASIFIKASI SENTIMEN PENGADUAN PUBLIK PADA FITUR LAKSA APLIKASI TANGERANG LIVE

Maura Davita Erzanda

ABSTRAK

Penelitian ini didorong oleh adanya asumsi internal di Diskominfo Kota Tangerang bahwa mayoritas laporan masyarakat pada fitur LAKSA berkonotasi negatif. Hingga kini, asumsi tersebut masih bersifat spekulatif karena belum dil.,98;./akukannya analisis data terkait untuk membuktikan asumsi internal tersebut. Penelitian ini bertujuan memberikan bukti empiris melalui analisis sentimen untuk memvalidasi asumsi tersebut secara objektif.

Menggunakan kerangka CRISP-DM, penelitian mengolah 10.586 data pengaduan periode 2021-2024. Tahap *preprocessing* meliputi *cleansing*, *stemming* (Sastrawi), dan *pseudo-labeling* ke dalam tiga kelas: positif, negatif, dan netral. Model yang dikembangkan membandingkan algoritma tunggal SVM dan *Random Forest* dengan model inovatif *Hybrid Stacking* (SVM-RF).

Hasil penelitian berhasil membuktikan asumsi internal bahwa mayoritas pengaduan masyarakat memang berkonotasi negatif, namun diikuti oleh proporsi sentimen netral yang signifikan. Model *Hybrid Stacking* menjadi solusi paling optimal dengan akurasi tertinggi sebesar 90,13%, melampaui performa model tunggal SVM (87,37%) dan *Random Forest* (83,71%). Penemuan ini merekomendasikan penggunaan model hibrida sebagai *Decision Support System* otomatis bagi Diskominfo untuk merespons keluhan publik dengan lebih presisi.

Kata kunci: Aplikasi Tangerang LIVE, Analisis Sentimen, Klasifikasi, Hibrida, Random Forest, SVM

**A HYBRID SVM-RANDOM FOREST MODEL TO OPTIMIZE PUBLIC
COMPLAINT SENTIMENT CLASSIFICATION ON THE LAKSA
FEATURE OF THE TANGERANG LIVE APPLICATION**

Maura Davita Erzanda

ABSTRACT (English)

This research is driven by an internal assumption at Diskominfo Tangerang City that the majority of public reports on the LAKSA feature are negative. To date, this assumption remains speculative. This study aims to provide empirical evidence through sentiment analysis to objectively validate this assumption.

Using the CRISP-DM framework, the study processed 10,586 complaint data points from 2021-2024. Preprocessing included cleansing, stemming (Sastrawi), and pseudo-labeling into three classes: positive, negative, and neutral. The models developed compared single SVM and Random Forest algorithms with an innovative Hybrid Stacking (SVM-RF) model.

The results successfully validate the internal assumption that the majority of public complaints carry a negative connotation, followed by a significant proportion of neutral sentiment. The Hybrid Stacking model emerged as the most optimal solution with the highest accuracy of 90.13%, outperforming the single SVM (87.37%) and Random Forest (83.71%) models. This finding recommends the hybrid model as an automated Decision Support System for Diskominfo to respond to public complaints more precisely

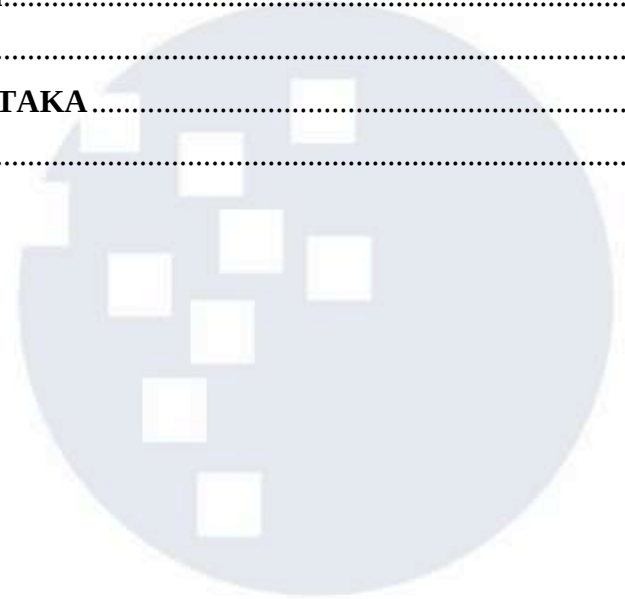
Keywords: *Classification, Tangerang LIVE app, Sentiment Analysis, Hybrid, Random Forest, SVM*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
LEMBAR PERNYATAAN PENGGUNAAN BANTUAN.....	iii
KECERDASAN ARTIFISIAL (AI)	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR.....	v
ABSTRAK	vii
<i>ABSTRACT (English)</i>	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR RUMUS	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah.....	6
1.4 Tujuan dan Manfaat Penelitian	7
1.4.1 Tujuan Penelitian	8
1.4.2 Manfaat Penelitian	9
1.5 Sistematika Penulisan	9
BAB II LANDASAN TEORI	12
2.1 Penelitian Terdahulu.....	12
2.2 Teori yang berkaitan.....	19
2.2.1 Teori <i>Smart City</i>	19
2.2.2 Teori <i>e-Government</i>	19
2.2.3 Teori <i>Super app</i>	20
2.2.4 Teori Analisis Sentimen	20
2.2.5 Teori Positif.....	20
2.2.6 Teori Negatif	21
2.2.7 Teori Netral.....	21

2.2.8	Teori <i>Natural Language Processing (NLP)</i> dan <i>Text Preprocessing</i>	22
2.2.9	Teori <i>Machine Learning</i>	22
1.2.7	Teori Hybrid	22
1.2.8	Teori <i>Confusion Matrix</i>	23
2.3	Framework/Algoritma yang digunakan.....	25
2.3.1	CRISP-DM	25
2.3.2	Random Forest.....	27
2.3.3	Support Vector Machine (SVM)	28
2.4	Tools/software yang digunakan	29
2.4.1	Python	29
2.4.2	Visual Studio Code	30
2.4.3	Jupyter Notebook.....	30
BAB III METODOLOGI PENELITIAN		32
3.1	Gambaran Umum Objek Penelitian.....	32
3.2	Metode Penelitian	33
3.2.1	Alur Penelitian.....	33
3.3	Teknik Pengumpulan Data.....	38
3.4	Teknik Analisis Data	38
3.4.1	<i>Data Preparation</i> dan <i>Data Processing</i>	39
3.4.2	<i>Text Preprocessing</i>	39
3.4.3	<i>Modeling</i>	39
3.4.4	Evaluasi Model.....	40
BAB IV ANALISIS DAN HASIL PENELITIAN		41
4.1	Business Understanding.....	41
4.2	Data Understanding	42
4.3	Data Preparation	47
4.3.1	Data Integration, Data Cleaning, Feature Engineering	47
4.3.2	Text Preprocessing.....	53
4.3.4	Data Labeling	66
4.4	Modeling.....	70
4.5	Evaluation	75

4.5.1	Random Forest.....	75
4.5.2	SVM.....	77
4.5.3	Hybrid SVM-RF	82
4.6	Deployment dan Hasil.....	85
BAB V SIMPULAN DAN SARAN.....		92
5.1	Simpulan.....	92
5.2	Saran.....	93
DAFTAR PUSTAKA.....		94
LAMPIRAN.....		103



UMN
 UNIVERSITAS
 MULTIMEDIA
 NUSANTARA

DAFTAR TABEL

Table 4.1 Kegunaan Library dan Modul	44
Table 4.2 Modeling Random Forest	73
Table 4.3 Perbandingan Kernel SVM	80
Table 4.4 Detail Konfigurasi Parameter.....	83
Table 4.5 Konfigurasi Meta-Classfier dan Stacking (Layer 1).....	84
Table 4.6 Comparison Table	85



DAFTAR GAMBAR

Gambar 1.1 Angka Pertumbuhan Aplikasi Tangerang LIVE	1
Gambar 1.2 langkah-langkah mengajukan pengaduan pada fitur LAKSA.....	2
Gambar 2.1 Metode CRISP-DM.....	25
Gambar 3.1 Alur Penelitian Sentimen Analisis	34
Gambar 4.1 Melihat 10 data teratas	42
Gambar 4.2 Data mentah.....	43
Gambar 4.3 Library	44
Gambar 4.4 Informasi dataset	46
Gambar 4.5 Informasi Dataset	46
Gambar 4.6 Pengubahan tipe data.....	47
Gambar 4.7 memfilter kolom yang dibutuhkan	48
Gambar 4.8 Pengecekan data duplikat	49
Gambar 4.9 Output pengecekan duplikasi data.....	49
Gambar 4.10 Pengecekan nilai none	50
Gambar 4.11 Output pengecekan nilai none	50
Gambar 4.12 Pengecekan nilai NaN	51
Gambar 4.13 Output pengecekan nilai NaN	52
Gambar 4.14 Menghapus nilai NaN	52
Gambar 4.15 Pendefinisian tanda baca	53
Gambar 4.16 Eksekusi pembersihan tanda baca	53
Gambar 4.17 lowercasing kolom isi_pengaduan	54
Gambar 4.18 Pustaka untuk tokenisasi	54
Gambar 4.19 Tokenisasi kata	55
Gambar 4.20 Pustaka untuk stopwords.....	56
Gambar 4.21 Daftar stopwords.....	56
Gambar 4.22 Eksekusi Stopword.....	57
Gambar 4.23 Customized stopwords.....	57
Gambar 4.24 Pustaka untuk WordNet	58
Gambar 4.25 Impor modul lematization	58
Gambar 4.26 Eksekusi lematization.....	59
Gambar 4.27 Contoh laporan setelah tahap word lemmatization	59
Gambar 4.28 Pustaka untuk Stemming.....	60
Gambar 4.29 Inisialisasi pustaka.....	60
Gambar 4.30 Pembuatan fungsi stemming	61
Gambar 4.31 Penerapan fungsi stemming	62
Gambar 4.32 Tampilan kolom isi_pengaduan	63
Gambar 4.33 Penghapusan duplikasi teks akhir	64

Gambar 4. 34 Output baris sebelum penghapusan duplikasi	64
Gambar 4.35 Output setelah penghapusan duplikasi	64
Gambar 4. 36 Jumlah laporan pengaduan berdasarkan tahun.....	65
Gambar 4. 37 Data labeling	66
Gambar 4.38 Pendefinisian kategori klasifikasi.....	67
Gambar 4.39 Proses eksekusi Pseudo-Labeling.....	68
Gambar 4.40 Rekonstruksi teks menjadi kalimat	69
Gambar 4.41 Pemisahan fitur (X) dan target (y).....	69
Gambar 4.42 Split dataset	70
Gambar 4.43 Vektorisasi teks dengan TF-IDF	71
Gambar 4.44 Inisialisasi dan Training SVM	72
Gambar 4.45 Inisialisasi dan Training Random Forest.....	72
Gambar 4.46 Confusion Matrix Random Forest.....	75
Gambar 4.47 Classification Report Random Forest	75
Gambar 4.48 Confusion Matrix SVM.....	77
Gambar 4.49 Classification Report SVM	78
Gambar 4.50 SVM : Linear.....	78
Gambar 4.51 SVM : RBF	79
Gambar 4.52 SVM : Polynomial.....	80
Gambar 4.53 Proses Hybrid SVM-RF	82
Gambar 4.54 Distribusi klasifikasi SVM berdasarkan tahun.....	86
Gambar 4.55 Distribusi klasifikasi random forest berdasarkan tahun	87
Gambar 4.56 Distribusi klasifikasi hybrid berdasarkan tahun	88
Gambar 4.57 Contoh Laporan yang sudah diberi label.....	88
Gambar 4.58 Frasa terbanyak muncul	90



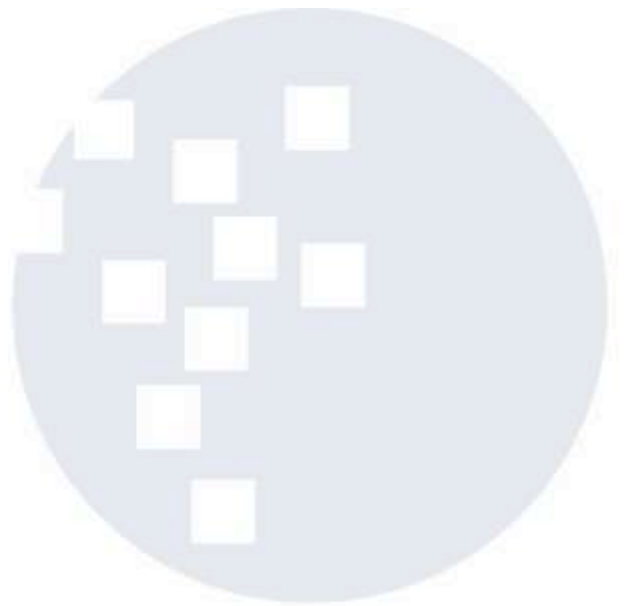
DAFTAR RUMUS

Rumus 2.1 Rumus Accuracy [44]	24
Rumus 2.2 Rumus Precision [44].....	24
Rumus 2.3 Rumus Recall [44]	25
Rumus 2.4 Rumus F1-Score [44].....	25
Rumus 2.5 Rumus Random Forest [48].....	27
Rumus 2.6 Rumus SVM [53].....	28



DAFTAR LAMPIRAN

Lampiran 1 Turnitin Similarity Report	103
Lampiran 2 Form Konsultasi Bimbingan.....	104
Lampiran 3 Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)	106



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA