

Deteksi Postingan Buzzer di Tiktok Menggunakan Model Klasifikasi Support Vector Machine (SVM) Berbasis Hibrida Fitur Stylometry

Esti Rahayu Putri*, Wina Witanti, Edvin Ramadhan

Universitas Jenderal Achmad Yani, Indonesia

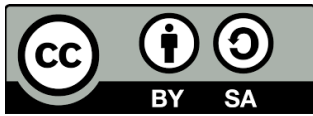
Email: estirahayuputri556@gmail.com*, witanti@gmail.com, edvin.ramadhan@gmail.com

Keywords:

buzzer, TikTok, stylometry, SVM, text classification

Abstract

The expansion of TikTok as a platform for public discussion has increased concerns about opinion manipulation by buzzer accounts, highlighting the need for computational detection of repetitive linguistic patterns. This study aimed to classify TikTok texts as buzzer or authentic netizen content using a Support Vector Machine (SVM) with hybrid stylometric and contextual features. A quantitative experimental approach was applied to 1,157 texts collected from TikTok. Preprocessing included text cleaning, tokenization, stemming, and Indonesian slang normalization. Stylometric features comprised Type-Token Ratio, Shannon entropy, and Hapax Legomena, combined with contextual features through feature concatenation. Classification employed an SVM with a Radial Basis Function kernel. The manuscript described fivefold cross-validation and a 92:8 training-testing split, comprising 1,066 training samples and 91 testing samples. Reported performance reached 94.44% accuracy, 97.30% precision, 96.00% recall, and 96.64% F1-score for the buzzer class. System analysis classified 72.3% of the dataset as buzzer content. These findings suggest that integrating linguistic diversity and contextual information can support the identification of potentially repetitive communication patterns. The approach offers a practical basis for automated TikTok text screening; however, classification results should be interpreted as indications of content resembling buzzers rather than definitive evidence of coordinated campaigns or verified account identities.



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright © 2026 by Author. Published by Internasional Journal Labs

Kata Kunci:

buzzer, TikTok, stylometry, SVM, klasifikasi teks

Abstrak

Perkembangan TikTok sebagai ruang diskusi publik meningkatkan kekhawatiran terhadap manipulasi opini oleh akun buzzer, sehingga diperlukan pendekatan komputasional untuk mendeteksi pola bahasa repetitif. Penelitian ini bertujuan mengklasifikasikan teks TikTok sebagai konten buzzer atau netizen autentik menggunakan Support Vector Machine (SVM) berbasis fitur hibrida stilometri dan kontekstual. Penelitian menggunakan pendekatan kuantitatif dengan desain eksperimental terhadap 1.157 teks yang dikumpulkan dari TikTok. Prapemrosesan meliputi pembersihan teks, tokenisasi, stemming, dan normalisasi bahasa gaul Indonesia. Fitur stilometri mencakup Type-Token Ratio, entropi Shannon, dan Hapax Legomena, yang digabungkan dengan fitur kontekstual melalui konkatenasi vektor. Klasifikasi menggunakan SVM dengan kernel Radial Basis Function. Naskah menjelaskan validasi silang lima lipatan dan pembagian data latih serta uji sebesar 92:8, terdiri atas 1.066 data latih dan 91 data uji. Kinerja yang dilaporkan mencapai akurasi 94,44%, presisi 97,30%, recall 96,00%, dan F1-score 96,64% untuk kelas buzzer. Analisis sistem mengklasifikasikan 72,3% dataset sebagai konten buzzer. Temuan tersebut menunjukkan bahwa integrasi keragaman bahasa dan informasi kontekstual berpotensi membantu identifikasi pola komunikasi repetitif. Pendekatan ini

menyediakan dasar praktis untuk penyaringan teks TikTok secara otomatis. Namun, hasil klasifikasi perlu dipahami sebagai indikasi karakteristik konten menyerupai buzzer, bukan bukti definitif kampanye terorganisasi maupun verifikasi identitas dan motivasi akun, sehingga interpretasinya harus mempertimbangkan batasan pelabelan serta konteks komunikasi.

PENDAHULUAN

(Duan et al., 2026; Akhtar et al., 2025) Perkembangan pesat media sosial, khususnya platform Tiktok, telah mengubah lanskap komunikasi dan diskusi publik di Indonesia. Media sosial kini menjadi arena utama pembentukan opini, terutama terkait isu-isu politik dan kebijakan strategis pemerintah. Namun, kemudahan akses ini juga membuka celah bagi fenomena manipulasi opini yang terstruktur, yang dilakukan oleh akun-akun *buzzer* terkoordinasi. Aktivitas *buzzer* ini yang dapat berupa *bot* otomatis maupun akun manusia bertujuan untuk mengamplifikasi narasi tertentu, meredam kritik, atau menciptakan polarisasi sentiment (Orabi et al., 2020).

Fenomena buzzer bukanlah hal yang sepenuhnya baru. Konsep menyebarkan informasi atau opini secara luas untuk mempengaruhi publik telah ada sejak lama dalam bentuk propaganda dan pemasaran dari mulut ke mulut. Namun, dengan pesatnya perkembangan teknologi dan media sosial, praktik ini mengalami transformasi signifikan (Wardana et al., 2025).

(Lingam & Das, 2025) Buzzer merupakan fenomena media sosial yang dapat diklasifikasikan ke dalam beberapa kategori utama berdasarkan motivasi dan tujuan komunikasinya. Secara umum, terdapat buzzer komersial yang berorientasi pada profit dengan membantu perusahaan meningkatkan visibilitas produk melalui ulasan positif dan kampanye tagar, serta buzzer politik yang bergerak demi kepentingan citra kandidat atau partai, terutama dalam upaya mendiskreditkan lawan maupun memengaruhi opini publik selama masa pemilihan (Li et al., 2025). Di sisi lain, terdapat pula buzzer isu sosial yang mendedikasikan aktivitasnya untuk advokasi kemanusiaan atau lingkungan hidup guna memobilisasi dukungan masyarakat terhadap gerakan tertentu. Selain kelompok yang terorganisir tersebut, muncul fenomena buzzer sukarela atau organik yang bergerak tanpa kompensasi materi, melainkan didorong oleh keyakinan pribadi sehingga pesan yang mereka sampaikan sering kali terasa lebih autentik bagi para pengikutnya. Memahami keragaman jenis buzzer ini sangat krusial agar kita dapat lebih kritis dalam menganalisis dampak serta etika di balik setiap informasi yang beredar di ruang digital (Amalia & Riki, 2024; Latah, 2020).

Cara kerja buzzer di media sosial merupakan sebuah proses sistematis yang mengandalkan koordinasi masif dan pemanfaatan teknologi untuk memengaruhi opini publik. Operasi ini dimulai dengan sinkronisasi pesan melalui arahan terpusat, di mana para buzzer menggunakan grup tertutup untuk mengatur waktu unggah serta penggunaan tagar agar sebuah topik dapat menjadi tren (Huang et al., 2025).

Dalam skala yang lebih besar, strategi ini sering kali diperkuat dengan penggunaan akun palsu atau bot guna menciptakan ilusi dukungan massal yang melampaui fakta di lapangan. Buzzer yang lebih canggih bahkan menerapkan teknik personalisasi dengan menganalisis data demografi audiens agar pesan yang disampaikan terasa lebih relevan dan

persuasif bagi target tertentu. Dengan taktik yang mencakup keterlibatan aktif dalam kolom komentar hingga penggunaan iklan terarah, buzzer mampu mengarahkan narasi digital secara efektif, sehingga menuntut pengguna media sosial untuk lebih kritis dalam memvalidasi setiap informasi yang mereka temui (Pal et al., 2023).

Dalam hal ini kita bisa melihat bahwa buzzer di soasial media itu, sangat mempengaruhi sudut pandang masyarakat luas. Semakin banyak buzzer yang bekerja hingga terus muncul di fyp netizen, akan semakin banyak yang mendukung opini tersebut. Menurut penulis kelemahan fitur tunggal itu sendiri hanya bisa melakukan analisis kosakata, tanpa bisa analisis akun buzzer atau netizen. Lalu kelemahan data di tiktok yaitu ada bahasa slang, gaul yang harus dibersihkan terlebih dahulu sebelum dilatih dan diuji (Hariyanto et al., 2024).

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimental untuk menguji efektifitas model klasifikasi *Support Vector Machine* (SVM) berbasis Hibrida fitur *Stylometry* dan juga menggunakan pendekatan NLP (*Natural Language Processing*) untuk menganalisa gaya kepenulisannya (Su & Jiang, 2025; Liu & Ma, 2025).

Fase Pengumpulan dan Pra-pemrosesan Data

Fase awal penelitian dimulai dengan Pengumpulan Data teks postingan (*Tiktok* utama, balasan, dan komentar) yang bersumber dari platform *Tiktok*. Data yang berhasil dikumpulkan selanjutnya melalui Pra-pemrosesan Teks yang ketat untuk memastikan kebersihan linguistik. Langkah-langkahnya meliputi *tokenisasi*, *stemming*, dan normalisasi bahasa *slang* Indonesia yang umum digunakan di *Tiktok*, sebelum digunakan untuk ekstraksi fitur (Nguyen et al., 2024).

Analisis pra-pemrosesan dalam penelitian ini tidak hanya sekedar membersihkan teks, tetapi juga bertujuan untuk mempertahankan karakteristik gaya bahasa (*stylistic features*) yang unik dari setiap penulis. Tahapan Normalisasi Bahasa Slang menjadi krusial karena buzzer di Indonesia seringkali menggunakan variasi penulisan yang sengaja disingkat atau diubah untuk menghindari deteksi algoritma standar. Dengan melakukan pemetaan (*mapping*) kata-kata tidak baku ke dalam bentuk formal, nilai frekuensi kata dasar dapat dihitung secara akurat, yang nantinya menjadi fondasi utama dalam penghitungan bobot TFIDF dan kekayaan kosakata pada tahap ekstraksi fitur.

Metode analisis hibrida yang menggabungkan fitur stilometri, struktural, dan algoritma *Support Vector Machine* (SVM) dirancang untuk mengatasi kompleksitas pola komunikasi buzzer yang semakin menyerupai manusia asli (Ellaky & Benabbou, 2024). Secara deskriptif, alasan penggunaan fitur stilometri adalah untuk membedah "sidik jari" linguistik dari teks ulasan, di mana sistem menghitung metrik seperti *Type-Token Ratio* (TTR), *Entropy*, dan *Hapax Legomena* guna mengidentifikasi kekayaan kosakata dan tingkat pengulangan kata yang sering menjadi ciri khas pesan terprogram atau repetitif. Penggabungan ini menjadi hibrida ketika sistem turut menyertakan fitur struktural untuk melihat pola teknis akun, sehingga deteksi tidak hanya bergantung pada apa yang dikatakan, tetapi juga bagaimana pesan tersebut disebarkan secara sistematis (Guyan et al., 2025).

Tahap *Advanced NLP Engine* menjadi fondasi krusial dalam metode ini karena bertugas membersihkan gangguan data melalui proses normalisasi bahasa gaul dan penyaringan kata henti, sehingga informasi yang masuk ke tahap ekstraksi fitur benar-benar

murni dan relevan. Data yang telah bertransformasi menjadi vektor hibrida ini kemudian diklasifikasikan menggunakan algoritma SVM dengan kernel *Radial Basis Function* (RBF) karena kemampuannya yang sangat efektif dalam memisahkan data teks berdimensi tinggi yang memiliki batas kelas tidak linier. Dengan pendekatan ini, admin dapat terus mengoptimalkan sistem melalui proses *retraining* model dan manajemen *ground truth*, yang secara keseluruhan menciptakan siklus deteksi cerdas yang adaptif terhadap perubahan tren bahasa maupun strategi buzzer di media sosial.

Akuisisi data dan Pelabelan (Grount Truth)

Kelas buzzer (Kelas 1)

Postingan yang berasal dari *cluster* akun dengan pola aktivitas ekstrem (volume postingan tinggi dalam waktu singkat) dan pola linguistik seragam (teks identik atau *paraphrase* yang berulang).

Kriteria netizen ontetik (Kelas 0)

Postingan yang berasal dari akun dengan pola aktivitas normal dan variasi linguistik yang tinggi.

Tahap ini menentukan target klasifikasi penelitian. Data sekunder dari Tiktok diberikan label biner untuk membedakan identitas akun secara objektif.

- Penjelasan: Objek data (D) dipetakan ke dalam label kelas y in $(0,1)$.
- Kelas 1 (Buzzer): Akun dengan aktivitas ekstrem dan pola teks serupa.
- Kelas 0 (Netizen): Akun dengan aktivitas normal dan gaya bahasa variatif.

Rumus Konsep: $D = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, di mana y_i adalah label kelas untuk data ke- i .

Klasifikasi Support Vector Machine (SVM)

SVM mencari pemisah (*hyperplane*) paling optimal dengan margin maksimal dalam ruang vektor hibrida.

Fungsi keputusan (Decision Function): sign (Fungsi Signum): Komponen ini berfungsi sebagai pengambil keputusan akhir. Jika nilai di dalam kurung positif, fungsi akan menghasilkan +1 (Buzzer), dan jika negatif akan menghasilkan -1 atau 0 (Netizen Otentik), $\sum_{i=1}^n$ (Summasi): Menunjukkan bahwa sistem menjumlahkan pengaruh dari seluruh *support vectors* (data latih yang paling dekat dengan garis pemisah) untuk mendapatkan hasil prediksi, a_i (Lagrange Multiplier): Parameter yang menentukan seberapa besar pengaruh atau "bobot" suatu data latih dalam membentuk garis pemisah. Data yang bukan merupakan *support vector* akan memiliki nilai nol, y_i (Label Kelas): Nilai target dari data latih, yaitu +1 untuk kategori buzzer dan 0/-1 untuk kategori netizen otentik. $K(x_i, x)$: Ini adalah fungsi kernel yang menghitung tingkat kemiripan antara data baru (x) dengan data yang sudah ada di database (x_i). b (Bias): Nilai konstanta yang menggeser posisi *hyperplane* agar mendapatkan pemisahan yang paling optimal di antara dua kelas data., lihat persamaan...9

$$= \text{sign} (x + 3.923)^n = \sum_{i=1}^n a_i y_i K(0.88, 0.11) + 1.0$$

$$f(x) = \text{sign} (x + a)^n = \sum_{i=1}^n a_i y_i K(x, x_i) + b$$

$f(x)$

$f(x)$

$$= \text{sign} (1 + 3.923)^n = \sum_{i=1}^n \text{aiyiK}(0.88,0.11) + 1.0$$

$$1 + 1.0 (3.923) \sum_{i=1}^n \text{aiyiK}(0.88,0.11)$$

Kernel RBF (Radial Basis Function):

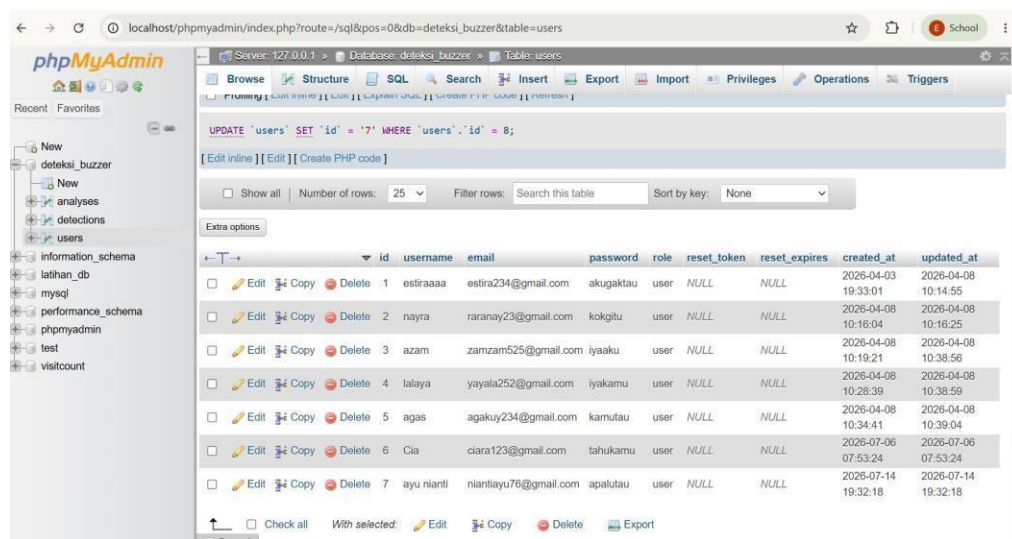
Digunakan untuk menangani pola perilaku yang tidak linear., lihat persamaa...10
 $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$

(Fungsi Kernel) Ini adalah hasil perhitungan kemiripan antara dua titik data, yaitu x_i (vektor fitur data latih) dan x_j (vektor fitur data lainnya), yang dipetakan ke dalam ruang dimensi tinggi agar lebih mudah dipisahkan oleh *hyperplane*, $\exp$ (Eksponensial) Fungsi basis alami ini memastikan bahwa nilai kemiripan akan meluruh secara halus saat jarak antar data semakin jauh, sehingga menghasilkan batas keputusan yang melengkung dan fleksibel, γ gamma (Parameter Gamma) Merupakan parameter penentu seberapa besar radius pengaruh dari satu titik data tunggal. nilai gamma yang diatur oleh Admin dalam menu "Optimasi Kernel RBF" akan menentukan seberapa sensitif model terhadap detail kecil dalam pola stilometri teks.

$$\|3.923 + (1 + 1.0)\| - 0.88.$$

Perancangan Tabel

Users



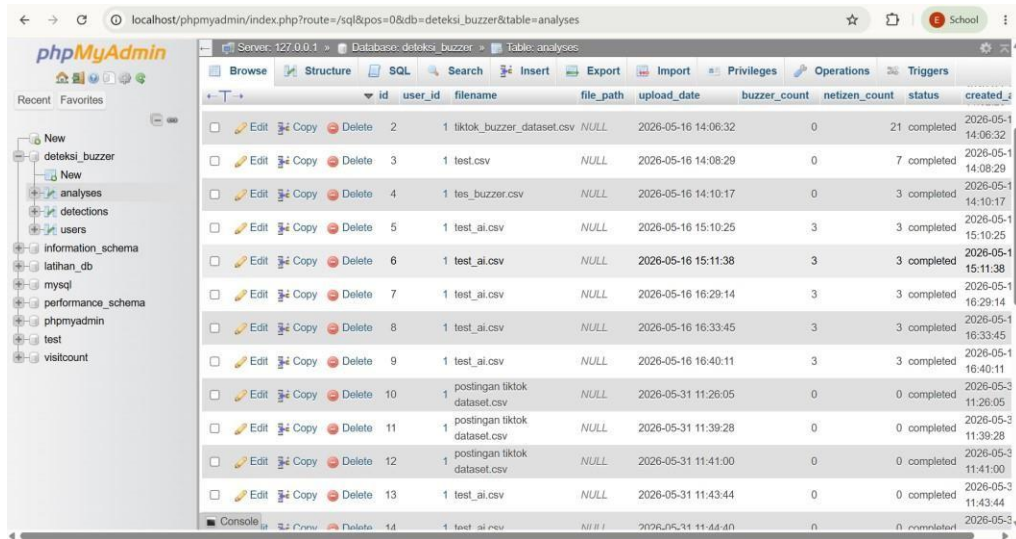
	id	username	email	password	role	reset_token	reset_expires	created_at	updated_at
☐	1	estiraaa	estira234@gmail.com	akugaktau	user	NULL	NULL	2026-04-03 19:33:01	2026-04-08 10:14:55
☐	2	nayra	aranay23@gmail.com	kokgitu	user	NULL	NULL	2026-04-08 10:16:04	2026-04-08 10:16:25
☐	3	azam	zamzam525@gmail.com	iyaaku	user	NULL	NULL	2026-04-08 10:19:21	2026-04-08 10:38:56
☐	4	lalaya	yayala252@gmail.com	iyakamu	user	NULL	NULL	2026-04-08 10:28:39	2026-04-08 10:38:59
☐	5	agas	agakuy234@gmail.com	kamutau	user	NULL	NULL	2026-04-08 10:34:41	2026-04-08 10:39:04
☐	6	Cia	ciara123@gmail.com	tahukamu	user	NULL	NULL	2026-07-06 07:53:24	2026-07-06 07:53:24
☐	7	ayu nianti	niantiaiyu76@gmail.com	apalutau	user	NULL	NULL	2026-07-14 19:32:18	2026-07-14 19:32:18

Gambar 2. Table users

Analisa

Analisa yang dapat saya simpulkan, bahwa ini merupakan table users, yang memiliki fungsi menyimpan database isinya akun seseorang yang telah melakukan login dan register.

Analyses



The screenshot shows the phpMyAdmin interface with the 'analyses' table selected. The table has the following columns: id, user_id, filename, file_path, upload_date, buzzer_count, netizen_count, status, and created_at. The data is as follows:

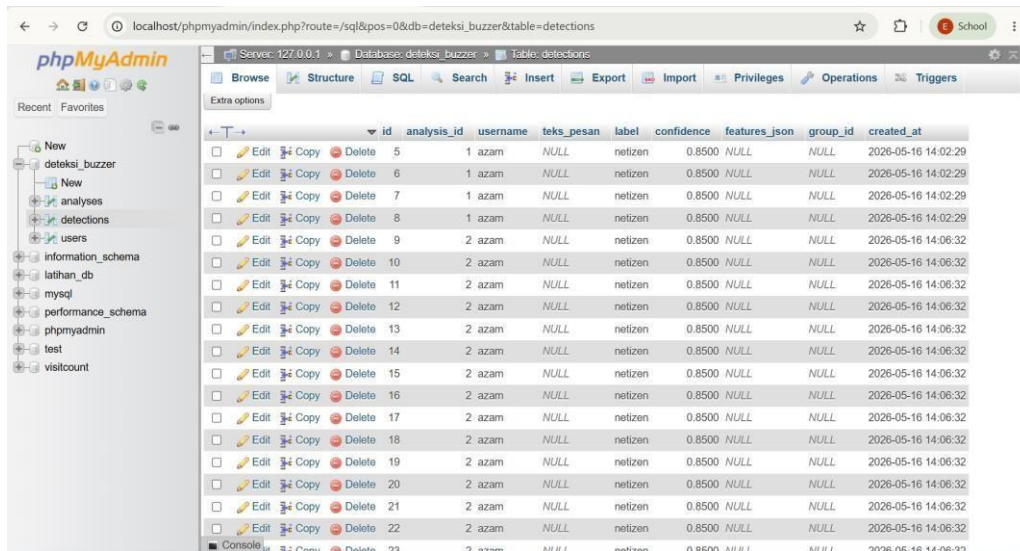
id	user_id	filename	file_path	upload_date	buzzer_count	netizen_count	status	created_at
2	1	tiktok_buzzer_dataset.csv	NULL	2026-05-16 14:06:32	0	21	completed	2026-05-16 14:06:32
3	1	test.csv	NULL	2026-05-16 14:08:29	0	7	completed	2026-05-16 14:08:29
4	1	tes_buzzer.csv	NULL	2026-05-16 14:10:17	0	3	completed	2026-05-16 14:10:17
5	1	test_ai.csv	NULL	2026-05-16 15:10:25	3	3	completed	2026-05-16 15:10:25
6	1	test_ai.csv	NULL	2026-05-16 15:11:38	3	3	completed	2026-05-16 15:11:38
7	1	test_ai.csv	NULL	2026-05-16 16:29:14	3	3	completed	2026-05-16 16:29:14
8	1	test_ai.csv	NULL	2026-05-16 16:33:45	3	3	completed	2026-05-16 16:33:45
9	1	test_ai.csv	NULL	2026-05-16 16:40:11	3	3	completed	2026-05-16 16:40:11
10	1	postingan tiktok dataset.csv	NULL	2026-05-31 11:26:05	0	0	completed	2026-05-31 11:26:05
11	1	postingan tiktok dataset.csv	NULL	2026-05-31 11:39:28	0	0	completed	2026-05-31 11:39:28
12	1	postingan tiktok dataset.csv	NULL	2026-05-31 11:41:00	0	0	completed	2026-05-31 11:41:00
13	1	test_ai.csv	NULL	2026-05-31 11:43:44	0	0	completed	2026-05-31 11:43:44

Gambar 3. Table Analyses

Analisa

Analisa yang dapat saya simpulkan, bahwa table ini merupakan table Analyses yang isi databasenya mengenai database yang menyimpan otomatis file csv yang di insert, data data ini nantinya akan melakukan proses untuk menentukan mana buzzer dan mana netizen otentik.

Detection



The screenshot shows the phpMyAdmin interface with the 'detections' table selected. The table has the following columns: id, analysis_id, username, teks_pesan, label, confidence, features_json, group_id, and created_at. The data is as follows:

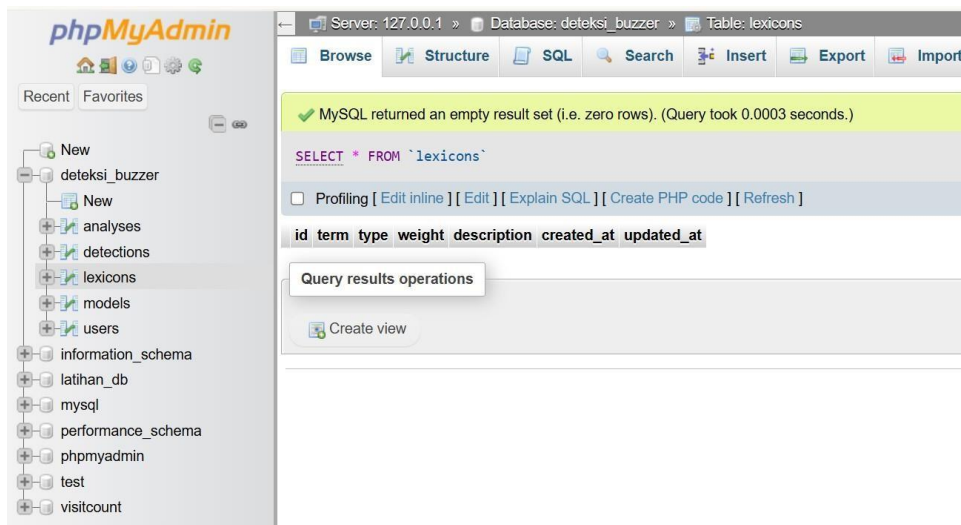
id	analysis_id	username	teks_pesan	label	confidence	features_json	group_id	created_at
5	1	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:02:29
6	1	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:02:29
7	1	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:02:29
8	1	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:02:29
9	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
10	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
11	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
12	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
13	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
14	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
15	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
16	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
17	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
18	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
19	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
20	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
21	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32
22	2	azam	NULL	netizen	0.8500	NULL	NULL	2026-05-16 14:06:32

Gambar 4. Table Detections

Analisa

Analisa yang dapat saya simpulkan, Berdasarkan tangkapan layar phpMyAdmin tersebut, tabel detections di dalam basis data deteksi_buzzer berfungsi sebagai tempat penyimpanan riwayat hasil deteksi/prediksi klasifikasi yang diproses oleh sistem.

Lexicons

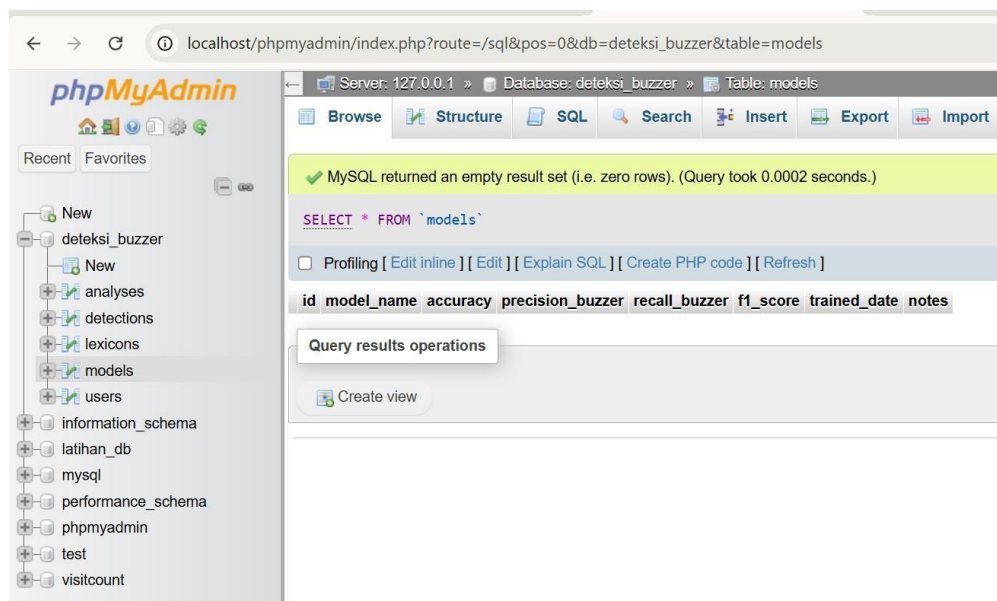


Gambar 5. Table Lexicons

Analisa

Gambar pertama menampilkan antarmuka pengelolaan basis data phpMyAdmin pada basis data deteksi_buzzer, secara spesifik menampilkan struktur entitas tabel lexicons. Tabel ini berfungsi sebagai repositori atau tempat penyimpanan kamus kata (*lexicon dictionary*) dan leksikon stilometri yang digunakan dalam tahap *Advanced NLP Preprocessing* maupun ekstraksi fitur. Struktur tabel ini dilengkapi dengan kolom-kolom penting seperti id sebagai *primary key*, term untuk menyimpan kata dasar/kata gaul (*slang*), type untuk mengategorikan jenis kata, weight untuk menyimpan nilai bobot leksikal, serta description beserta timestamp (*created_at*, *updated_at*).

Models



Gambar 6. Table Models

Analisa

Gambar kedua memperlihatkan antarmuka phpMyAdmin untuk tabel models di dalam basis data deteksi_buzzer. Tabel ini berfungsi khusus sebagai wadah penyimpanan riwayat (*history log*), parameter, dan hasil evaluasi performa dari model pembelajaran mesin Support Vector Machine (SVM) yang telah dilatih (*retrained*). Struktur kolom pada tabel ini mencakup id, model_name (nama/versi model), accuracy, precision_buzzer, recall_buzzer, f1_score, trained_date (waktu pelatihan), dan notes.

Penjelasan Desain antarmuka Buzzer

Koneksi

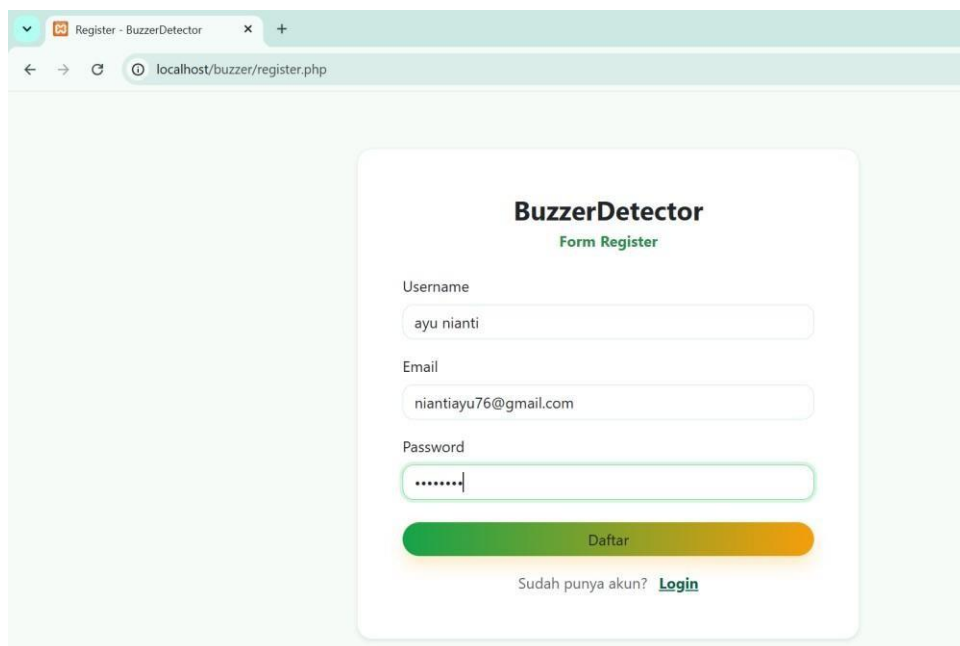
Dapat dilihat dari source code di atas bahwa source code ini merupakan koneksi database, yang nantinya merupakan patokan database untuk menyambung ke mysql. Koneksi ini juga akan dipanggil ke page yang lainnya.

Deteksi.sql

Dari source yang dilihat dari atas, dapat dianalisa bahwa source tersebut adalah sekumpulan database sql mengenai deteksi buzzer, source code itu nantinya menjadi wadah database untuk menampung data yang dijalankan di website.

Halaman Register

Tampilan Desain

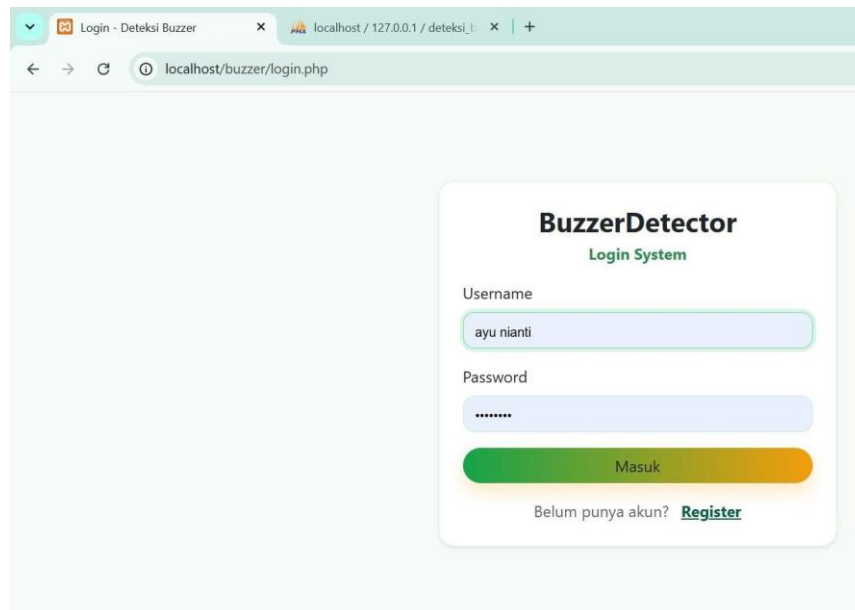
A screenshot of a web browser showing a registration form titled "BuzzerDetector Form Register". The form is centered on a light green background. It contains three input fields: "Username" with the value "ayu nianti", "Email" with the value "niantiayu76@gmail.com", and "Password" with masked characters ".....". Below the fields is a large green button labeled "Daftar". At the bottom, there is a link that says "Sudah punya akun? Login". The browser's address bar shows "localhost/buzzer/register.php".

Gambar 7. Proses register deteksi buzzer

Analisa yang saya dapatkan dari page register ini yaitu, source code ini melakukan sebuah insert register yang nantinya, akan nyambung ke login, jika register tersebut berhasil dilakukan.

Halaman Login

Tampilan Desain



The screenshot shows a web browser window with the address bar displaying 'localhost/buzzer/login.php'. The page features a login form titled 'BuzzerDetector Login System'. The form includes a 'Username' field with the text 'ayu nianti', a 'Password' field with masked characters '*****', and a green 'Masuk' button. Below the button is a link that says 'Belum punya akun? [Register](#)'.

Gambar 8. proses login deteksi buzzer

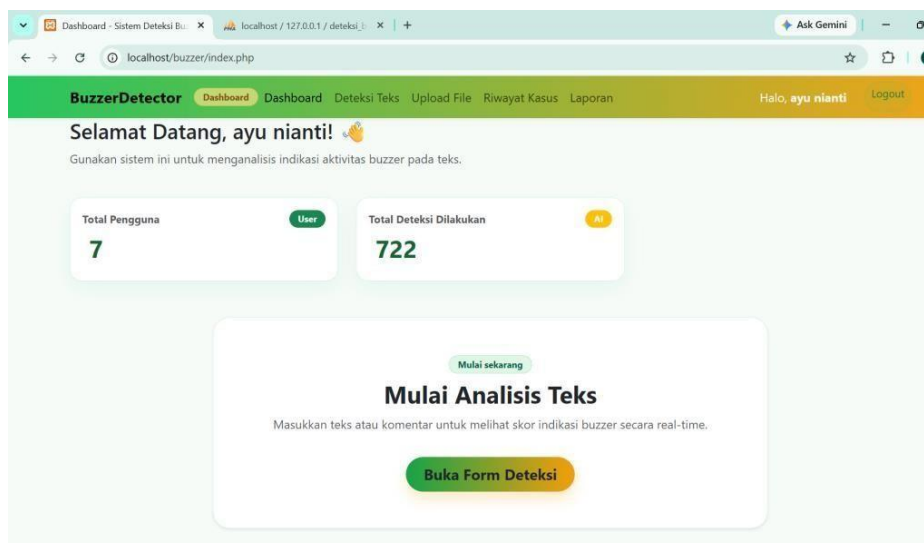
Analisa yang dapat saya simpulkan di page login ini adalah, memanggil dataset yang sudah diinsert di register, data tersebut kembali dimasukkan untuk melakukan proses login, jika datanya sesuai dengan apa yang dilakukan di register, maka proses login berhasil.

Halaman Logout

Setelah sesi dibersihkan, kode ini langsung mengalihkan (*redirect*) halaman pengguna ke login.php dan menghentikan eksekusi skrip agar tidak ada kode di bawahnya yang berjalan.

Halaman Dashboard

Tampilan Desain



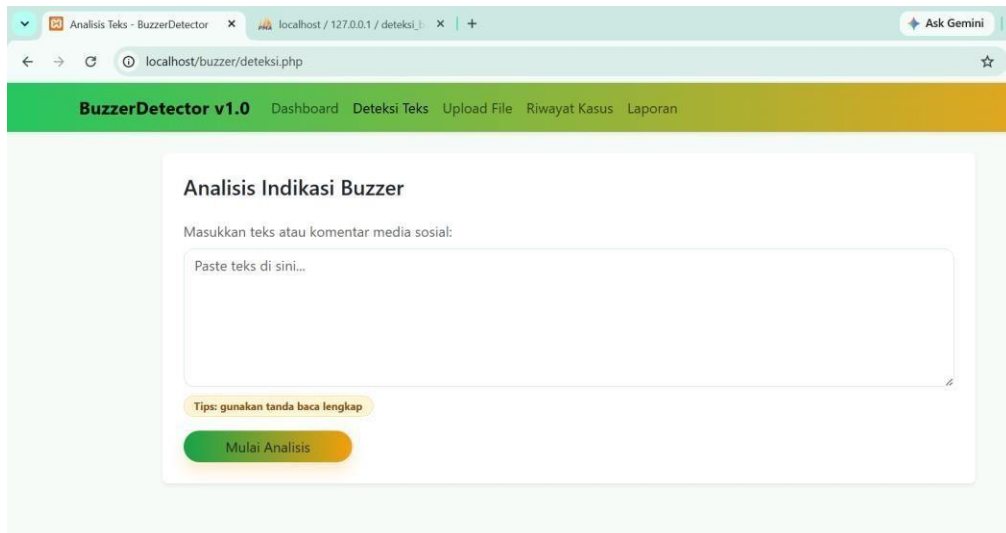
The screenshot shows the dashboard of the BuzzerDetector system. The top navigation bar includes links for 'Dashboard', 'Deteksi Teks', 'Upload File', 'Riwayat Kasus', and 'Laporan'. The user is logged in as 'ayu nianti'. The main content area displays a welcome message 'Selamat Datang, ayu nianti!' and a brief description of the system's purpose. Two summary cards show 'Total Pengguna' (7) and 'Total Deteksi Dilakukan' (722). A central section titled 'Mulai Analisis Teks' encourages users to input text for real-time analysis, with a 'Buka Form Deteksi' button.

Gambar 9. dashboard deteksi buzzer

Analisa yang dapat saya simpulkan ini adalah sebuah halaman dashboard, di sini dapat dilihat bahwa sebanyak data 722 yang telah diuji, dan terdapat 7 pengguna yang sudah melakukan registrasi di website ini. Lalu, di bawahnya ada tools tentang mulai analisa teks, maka tinggal klik saja form tersebut untuk melakukan analisa teks.

Halaman Deteksi

Tampilan Desain

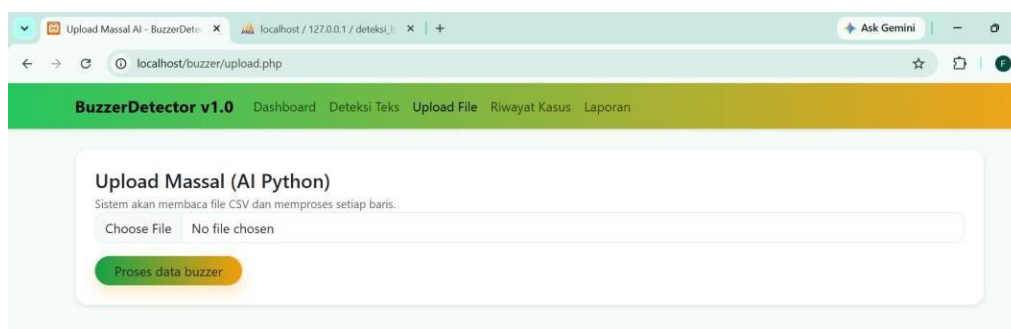


Gambar 10. Deteksi teks buzzer

Analisa yang dapat saya simpulkan, bahwa ini merupakan sebuah page untuk menganalisa sebuah teks yang telah ditentukan, untuk menganalisa apakah ini adalah buzzer atau netizen otentik.

Upload

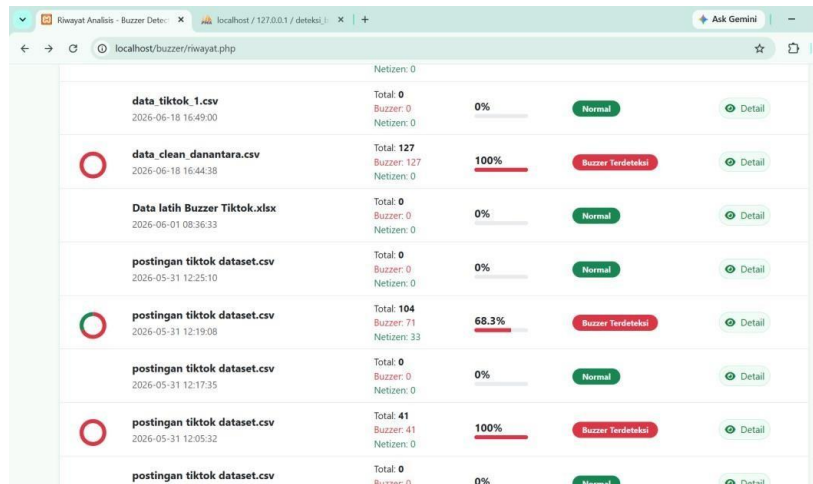
Tampilan Desain



Gambar 11. upload file deteksi buzzer

Analisa yang dapat saya simpulkan dari page upload ini, yaitu ini fungsinya untuk mengupload sebuah file csv. Yang nantinya di proses file csv tersebut, berapa buzzer yang ada dan berapa netizen otentik yang ada di page tersebut.

Halaman Riwayat Tampilan Desain



File Name	Total	Buzzer	Netizen	Status	Action
data_tiktok_1.csv 2026-06-18 16:49:00	0	0	0	Normal	Detail
data_clean_danantara.csv 2026-06-18 16:44:38	127	127	0	Buzzer terdeteksi	Detail
Data latih Buzzer Tiktok.xlsx 2026-06-01 08:36:33	0	0	0	Normal	Detail
postingan tiktok dataset.csv 2026-05-31 12:25:10	0	0	0	Normal	Detail
postingan tiktok dataset.csv 2026-05-31 12:19:08	104	71	33	Buzzer terdeteksi	Detail
postingan tiktok dataset.csv 2026-05-31 12:17:35	0	0	0	Normal	Detail
postingan tiktok dataset.csv 2026-05-31 12:05:32	41	41	0	Buzzer terdeteksi	Detail
postingan tiktok dataset.csv	0	0	0	Normal	Detail

Gambar 12. riwayat laporan deteksi buzzer



Gambar 13. detail riwayat laporan

Analisa yang dapat saya simpulkan yaitu, seperti dijelaskan tadi hasil dari analisa form deteksi dan juga analisa dari upload file csv semuanya ada di riwayat laporan, jika seseorang ingin melihat secara detail maka akan menampilkan laporan yang degtail berapa persentase buzzer dan netizen otentik.

HASIL DAN PEMBAHASAN

Referensi literatur terkait

Analisis terhadap fenomena *buzzer* di media digital memerlukan pendekatan komputasional yang mampu mengidentifikasi pola kepenulisan unik dari aktor-aktor di balik akun media sosial (Modupe et al., 2022). Secara teoretis, identifikasi ini bertumpu pada konsep *authorship attribution* berbasis *stylometry*, yaitu teknik yang mengeksplorasi gaya bahasa spesifik penulis untuk mengenali identitas atau karakteristik teks. Keandalan metode ini ditunjukkan oleh Abbass dan Al-Zubaidi (2021a, 2021b) serta Vasilakis dan Maragoudakis (2021) yang membuktikan bahwa klasifikasi *stylometric* multi-kelas mampu

mengenali pola kepenulisan secara presisi. Kemampuan ekstraksi fitur gaya bahasa ini diperkuat oleh Prasath dan O'Rior dan (2021) melalui penerapan *Natural Language Processing* (NLP), yang didukung oleh temuan Lukin doting (2023) mengenai efektivitas penggunaan adjektiva dan adverbial sebagai parameter kunci dalam memperjelas karakter dan corak tulisan yang unik (Raja & Raj, 2022).

Dalam perkembangannya, algoritma *Support Vector Machine* (SVM) menjadi salah satu pilar utama yang paling konsisten digunakan untuk mengklasifikasikan fitur-fitur *stylometric* tersebut. Keunggulan performa SVM dalam memetakan batas klasifikasi teks yang kompleks telah diteliti secara mendalam oleh Cercas del Pozas dan Li (2020), Bavani dan Asaduzzaman (2022), serta Tariq dkk. (2021) yang membandingkannya dengan *Random Forest* untuk atribusi skala besar. Kestabilan performa SVM ini sangat bergantung pada pemilihan kernel dan optimasi fitur, sebagaimana dibuktikan oleh Kumar dan Sharma (2023) terkait ketahanan kernel SVM pada dokumen teknis, serta Dhar dan Dasgupta (2023) yang berfokus pada peningkatan seleksi fitur untuk forensik *stylometric*. Kompleksitas ini semakin disempurnakan oleh Gomez-Adorno dkk. (2020) yang memperluas implementasi SVM untuk kebutuhan *cross-domain authorship attribution*, menjadikannya sangat relevan untuk melacak pergerakan *buzzer* yang sering kali beroperasi lintas platform digital.

Pada spektrum yang lebih luas, keterkaitan antara *stylometry* dan manipulasi teks juga merambah pada aspek analisis forensik dan adaptasi model terhadap kecerdasan buatan generatif. Kajian mengenai *code authorship identification* oleh Wang dan Chen (2022) memperlihatkan bahwa analisis gaya tidak hanya terbatas pada teks opini, melainkan juga pada struktur logika pemrograman digital.

Kebutuhan forensik ini sejalan dengan penelitian Rz. dan Badieah (2025) yang sukses menerapkan *stylometry* dan SVM untuk deteksi plagiarisme teks, serta Aggazzotti dan Smith (2025) yang memperluas domainnya pada transkrip wicara.

Menghadapi tantangan manipulasi teks modern, riset dari Sari dan Al Faridzi (2023) mengenai *authorship obfuscation* dalam Bahasa Indonesia, serta analisis Kumarage dan Liu (2023) terhadap *Large Language Models* (LLM), menegaskan bahwa pemahaman mendalam mengenai literatur *stylometric* ini sangat krusial untuk membongkar taktik kamuflase teks yang kerap digunakan oleh jaringan *buzzer* masa kini.

Desain Eksperimen & Skenario Pembagian Data

Penelitian ini menggunakan total 1157 data, yang dikumpulkan dari media sosial tiktok. Data dibagi menggunakan metode k-Fold Cross-Validation (k=5) dan pembagian Train-Test split (92:8). Data latih sebanyak 1066 data sedangkan data latih sebanyak 91 data uji.

Tabel 2. Dataset *no clean* dan *clean*

No	No Clean	Clean
1	Kemarin dapet susu yg expired dong di sekolahan anak gw, tolong dievaluasi ketat dong kateringnya!!	kemarin dapat susu yang expired dong di sekolahan anak saya tolong dievaluasi ketat dong kateringnya
2	Sangat terbantu dg program mbg ini, anak saya jd dpt makan sehat gratis tiap hari, makasih pak prabowo	sangat terbantu dengan program mbg ini anak saya jadi dapat makan sehat gratis tiap hari makasih pak prabowo

No	No Clean	Clean
3	Penasaran sama anggaran rincian per porsi makanan mbg ini berapa sih sebenarnya?	penasaran sama anggaran rincian per porsi makanan mbg ini berapa sih sebenarnya
4	Kebijakan pencitraan doang nih, utang negara bengkak cm buat makan siang, sedih bgt liatnya	kebijakan pencitraan doang nih utang negara bengkak cuma buat makan siang sedih banget liatnya
5	Kemarin dapet susu yg expired dong di sekolahan anak gw, tolong dievaluasi ketat dong kateringnya!!	kemarin dapat susu yang expired dong di sekolahan anak saya tolong dievaluasi ketat dong kateringnya
6	Kira-kira menunya disamakan se-Indonesia atau beda-beda tiap daerah ya?	kirakira menunya disamakan seindonesia atau bedabeda tiap daerah ya
7	Ini program pemborosan anggaran negara aja, utang makin numpuk gara2 mbg ckckck	ini program pemborosan anggaran negara aja utang makin numpuk gara mbg ckckck
8	Anggarannya gede bgt mending buat perbaiki fasilitas sekolah yg rusak drpd mbg begini	anggarannya gede banget mending buat perbaiki fasilitas sekolah yang rusak drpd mbg begini
9	Kita pantau aja dulu kembangannya, semoga beneran manfaat buat adek-adek sekolah	kita pantau aja dulu perkembangannya semoga beneran bermanfaat buat adekadek sekolah
10	Mantap mbgg!! Gizi anak indonesia makin terjamin, ga kasian lagi liat anak kelaparan di kls	mantap mbgg gizi anak indonesia makin terjamin tidak kasian lagi liat anak kelaparan di kls

Ekstraksi Fitur Hibrida (Stilometri & Kontekstual)

Metode hibrida menggabungkan "cara menulis" (Stilometri) dan "makna kata" (Kontekstual) ke dalam satu vektor fitur.

Type-Token Ratio (TTR): Mengukur kekayaan kosakata. Buzzer cenderung memiliki skor rendah karena repetisi., lihat persamaan...5 kemarin dapat susu yang expired dong di sekolahan anak saya tolong dievaluasi ketat dong kateringnya

$TTR = V/N$

Tokenization (Menghitung $N(T)$)

Ini sama sekali tidak benar, jika saya memaksanya untuk menonton semua video di FYP, bersama saya.

- 1) Kemarin
- 2) dapat
- 3) susu
- 4) yang
- 5) Expired
- 6) dong
- 7) di
- 8) sekolahan
- 9) anak
- 10) saya
- 11) tolong
- 12) Di
- 13) evaluasi
- 14) ketat
- 15) dong

16) cateringnya Total $N(T) = 16$

Mencari Kata Unik (Menghitung $V(T)$)

Identifikasi kata yang berbeda. Karena kelima kata di atas berbeda semua, maka: Total $V(T) = 16$.

Perhitungan

$$TTR = \frac{16}{16} = 1.0$$

(Skor 1.0 berarti variasi kata sangat sempurna/tidak ada pengulangan).

Shannon Entropy: Mengukur kompleksitas informasi teks untuk mendeteksi pesan hasil *copy-paste*., lihat persamaan...6

$$H(X) = - \sum_{i=1}^n P(x_i) \log_2 P(x_i)$$

Tentukan Peluang $P(x_i)$: Karena di teksmu semua kata unik (tidak ada pengulangan), maka setiap kata muncul 1 kali dari total 16 kata. Jadi, peluang setiap kata adalah $P(x_i)$

$$= \frac{1}{16}$$

Masukkan ke Rumus Karena ada 16 kata yang peluangnya sama, kita bisa menyederhanakan rumusnya menjadi

$$H(X) = -16 \times \frac{1}{16} \times \log_2 \frac{1}{16}$$

$\log_2 (1/16)$ adalah sekitar -3.923

$$H(X) = (5 \times 1/16)$$

Hapax Legomena: Mengukur orisinalitas melalui kata yang hanya muncul tepat satu kali., lihat persamaan...7

$$HL = \frac{V_1}{N}$$

Total kata $N(16)$

$$HL = \frac{16}{16} = 1.0$$

Karena semua kata unik dan tidak ada yang diulang, maka semua kata tersebut masuk kategori muncul 1 kali. Jadi, $V_1 = 16$.

Integrasi Vektor Fitur (Concatenation): Menggabungkan vektor stilometri (F_{sty}) dan kontekstual (F_{kon}) menjadi satu., lihat persamaan...8

$$V_{hibrida} = [F_{sty} + F_{kon}]$$

Tentukan isi vektor Stilometri (F_{sty})

Vektor ini berisi fitur-fitur statistik yang sudah kamu hitung tadi: $TTR = 1.0$

Entropy = 3.923 Hapax Legomena = 1.0

$$Maka: F_{sty} = [1.0, 3.923, 1.0]$$

Vektor ini biasanya dihasilkan dari model bahasa (seperti Word2Vec). Misalkan model memberikan nilai bobot kata:

$$Maka: F_{kon} = [0.88, 0.11].$$

Kamu cukup menyatukan angka-angka tersebut ke dalam satu baris: $V_{hibrida} = [1.0, 3.923, 1.0, 0.88, 0.11]$.

Klasifikasi Support Vector Machine (SVM)

SVM mencari pemisah (*hyperplane*) paling optimal dengan margin maksimal dalam ruang vektor hibrida.

Fungsi keputusan (Decision Function): sign (Fungsi Signum): Komponen ini berfungsi sebagai pengambil keputusan akhir. Jika nilai di dalam kurung positif, fungsi akan menghasilkan +1 (Buzzer), dan jika negatif akan menghasilkan -1 atau 0 (Netizen Otentik), $\sum_{i=1}^n$ (Summation): Menunjukkan bahwa sistem menjumlahkan pengaruh dari seluruh *support vectors* (data latih yang paling dekat dengan garis pemisah) untuk mendapatkan hasil prediksi, a_i (Lagrange Multiplier): Parameter yang menentukan seberapa besar pengaruh atau "bobot" suatu data latih dalam membentuk garis pemisah. Data yang bukan merupakan *support vector* akan memiliki nilai a_i nol, y_i (Label Kelas): Nilai target dari data latih, yaitu +1 untuk kategori buzzer dan 0/-1 untuk kategori netizen otentik. $K(x_i, x)$: Ini adalah fungsi kernel yang menghitung tingkat kemiripan antara data baru (x) dengan data yang sudah ada di database (x_i). b (Bias): Nilai konstanta yang menggeser posisi *hyperplane* agar mendapatkan pemisahan yang paling optimal di antara dua kelas data.

lihat persamaan...9 $f(x) = \text{sign}(x + a)^n = \sum_{i=1}^n i=1$

$$f(x) = \text{sign}(x + 3.923)^n = \sum_{i=1}^n i=1$$

$$f(x) = \text{sign}(1 + 3.923)^n = \sum_{i=1}^n i=1$$

$$a_i y_i K(0.88, 0.11) + 1.0$$

$$a_i y_i K(0.88, 0.11) + 1.0$$

$$1 + 1.0 (3.923) \sum_{i=1}^n$$

$$a_i y_i K(0.88, 0.11)$$

Kernel RBF (Radial Basis Function)

Digunakan untuk menangani pola perilaku yang tidak linear., lihat persamaa...10 $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$

(Fungsi Kernel) Ini adalah hasil perhitungan kemiripan antara dua titik data, yaitu x_i (vektor fitur data latih) dan x_j (vektor fitur data lainnya), yang dipetakan ke dalam ruang dimensi tinggi agar lebih mudah dipisahkan oleh *hyperplane*, $\exp$ (Eksponensial):

Fungsi basis alami ini memastikan bahwa nilai kemiripan akan meluruh secara halus saat jarak antar data semakin jauh, sehingga menghasilkan batas keputusan yang melengkung dan fleksibel, $-\gamma$ gamma (Parameter Gamma): Merupakan parameter penentu seberapa besar radius pengaruh dari satu titik data tunggal; nilai gamma yang diatur oleh Admin dalam menu "Optimasi Kernel RBF" akan menentukan seberapa sensitif model terhadap detail kecil dalam pola stilometri teks.

$\|x_i - x_j\|^2$ (Squared Euclidean Distance): Komponen ini menghitung kuadrat jarak geometris antara dua vektor fitur hibrida (gabungan skor Stilometri dan Struktural); semakin kecil jarak ini, berarti kedua ulasan tersebut memiliki karakteristik gaya penulisan dan pola interaksi yang sangat mirip.

$$\|3.923 + (1 + 1.0)\| - 0.88.$$

Pelabelan data scrapping menggunakan program

Pipeline_cleaning_manual.py

Fungsi `bersihkan_teks_manual` bekerja sebagai mesin pembersih utama yang memproses teks baris demi baris menggunakan metode manipulasi string dan *Regular Expression* (Regex). Proses pembersihan ini meliputi pengubahan teks menjadi huruf kecil (*case folding*), penghapusan komponen eksternal (seperti tautan URL, tagar #, dan sebutan akun @), eliminasi simbol/emoji non-ASCII, serta pembuangan angka dan tanda baca.

Pelabelan menggunakan Program

Deteksi Teks

Ketika teks dikirimkan dari formulir PHP, data tersebut tidak langsung diolah di server web, melainkan dikonversi ke format universal JSON agar dapat dipahami oleh ekosistem Python yang menjalankan model *Natural Language Processing* (NLP). Pemisahan tanggung jawab (*separation of concerns*) ini sangat efisien karena PHP fokus menangani manajemen sesi, interaksi pengguna, dan penyimpanan basis data, sementara proses komputasi berat untuk klasifikasi *buzzer* sepenuhnya diserahkan kepada Python.

Alur ini ditutup dengan sinkronisasi data kembali ke database setelah API memberikan umpan balik, memastikan sistem tetap responsif dan data tersimpan secara konsisten.

Upload teks

Alur pemrosesan massal (*bulk processing*) ini menggambarkan bagaimana sistem web PHP mengelola beban kerja berat (*heavy workload*) secara sekuensial dengan membagi beban kerja tersebut ke dalam siklus yang terstruktur. Dibandingkan mengirimkan seluruh data CSV sekaligus yang berisiko membuat memori server kelebihan beban, PHP memilih pendekatan bertahap dengan membaca, mengirimkan ke API AI Python, dan menyimpan hasilnya baris demi baris (*row-by-row*).

Metode ini sangat efisien karena menjaga penggunaan memori tetap rendah dan memungkinkan adanya penanganan kesalahan (*error handling*) yang cepat; jika API AI terputus di tengah jalan, sistem dapat langsung menghentikan proses untuk mencegah kerusakan data lebih lanjut. Di akhir siklus, integrasi data ditutup dengan memperbarui ringkasan statistik pada tabel induk, mengubah proses yang kompleks menjadi satu laporan final yang rapi dan informatif bagi pengguna.

Riwayat Laporan dan detail

Tampilan desain



Gambar 14. Hasil analisa buzzer

Saat halaman dimuat, PHP langsung menghitung jumlah baris data (*row count*) di tabel `detections` untuk masing-masing label, yaitu 'buzzer' dan 'netizen'. Dari angka tersebut,

PHP melakukan kalkulasi persentase untuk mencari tahu seberapa besar dominasi akun *buzzer* di dalam ekosistem digital yang sedang diaudit.

Analisis Ambang Batas (Thresholding) Sistem menerapkan logika keputusan sederhana berbasis persentase. Jika proporsi konten *buzzer* yang terdeteksi melampaui angka 30% dari total data, sistem secara otomatis menganalisis dan menyimpulkan situasi tersebut sebagai "indikasi kuat adanya kampanye terorganisir (manipulasi opini)". Sebaliknya, jika di bawah itu, sistem menilainya sebagai opini organik sehat dari netizen biasa.

Berdasarkan analisis algoritma, ditemukan tingkat penetrasi *buzzer* sebesar 72.3%. Terdapat indikasi kuat adanya kampanye terorganisir pada dataset ini. Dari analisa ini kita bisa menarik kesimpulan, bahwa dataset yang berisi data 1157 itu di analisa oleh website buatan penulis untuk mengetahui dataset ini *buzzer* dan netizen otentiknya ada berapa. Setelah proses analisa selesai, maka akan tinggal masuk ke halaman Laporan untuk cetak laporan berupa pdf. Dominasi jumlah data *buzzer* dapat mengindikasikan bahwa aktivitas *buzzer* di platform media sosial yang menjadi objek penelitian cukup tinggi.

KESIMPULAN

Kesimpulan yang dapat saya simpulkan dari metode yang telah saya ambil untuk penelitian ini yaitu, saya telah menetapkan dua metode campuran untuk menyelesaikan permasalahan penelitian ini. SVM sendiri ditentukan untuk mendalami pemisahan permasalahan netizen otentik dan *buzzer*. Yang seperti diketahui sebelumnya metode *stylometri* ini hanya untuk menganalisa kasus dataset yang panjang, maka dari itu saya memilih platfrom tiktok yang jelas, memiliki karakter yang sangat panjang untuk menjelaskan sesuatu di postingan platfrom tiktok (Lukin et al., 2023; (Rini & & Prawiranegara, 2024).

Berdasarkan analisis stilometrik terhadap teks sampel, proses ekstraksi fitur dimulai dengan tahap tokenisasi yang menghasilkan total $SN = 16$ kata (token). Uji Type-Token Ratio (TTR) menghasilkan nilai 1,0, yang menunjukkan variasi kosakata maksimum karena seluruh kata teridentifikasi sebagai kata unik ($V = 16$). Komposisi ini selaras dengan nilai Hapax Legomena (HL) sebesar 1,0, di mana 100% kata dalam teks tersebut muncul tepat satu kali. Nilai Entropi Shannon sebesar 3,923 bit mengonfirmasi tingkat kompleksitas dan ketidakpastian informasi yang tinggi, yang mencerminkan pola teks alami tanpa pengulangan (copy-paste). Selanjutnya, ketiga fitur statistik ini dibentuk menjadi vektor stilometrik, $F_{\text{sty}} = [1,0, 3,923, 1,0]$. Melalui teknik konkatenasi, F_{sty} diintegrasikan dengan vektor kontekstual berbasis pembobotan bahasa, $F_{\text{kon}} = [0,88, 0,11]$, sehingga menghasilkan vektor fitur hibrida, $V_{\text{hibrida}} = [1,0, 3,923, 1,0, 0,88, 0,11]$, yang siap untuk tahap klasifikasi..

Kernel RBF (Radial Basis Function) Digunakan untuk menangani pola perilaku yang tidak linear., lihat persamaa...10 $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ (Fungsi Kernel) Ini adalah hasil perhitungan kemiripan antara dua titik data, yaitu x_i (vektor fitur data latih) dan x_j (vektor fitur data lainnya), yang dipetakan ke dalam ruang dimensi tinggi agar lebih mudah dipisahkan oleh *hyperplane*, $\exp$ (Ekspensial): Fungsi basis alami ini memastikan bahwa nilai kemiripan akan meluruh secara halus saat jarak antar data semakin jauh, sehingga menghasilkan batas keputusan yang melengkung dan fleksibel, γ (Parameter Gamma): Merupakan parameter penentu seberapa besar radius pengaruh dari satu titik data

tunggal; nilai gamma yang diatur oleh Admin dalam menu "Optimasi Kernel RBF" akan menentukan seberapa sensitif model terhadap detail kecil dalam pola stilometri teks. $\|x_i - x_j\|_2$ (Squared Euclidean Distance): Komponen ini menghitung kuadrat jarak geometris antara dua vektor fitur hibrida (gabungan skor Stilometri dan Struktural); semakin kecil jarak ini, berarti kedua ulasan tersebut memiliki karakteristik gaya penulisan dan pola interaksi yang sangat mirip. $\|3.923 + (1 + 1.0)\| - 0.88$.

Kinerja model SVM menunjukkan performa yang sangat baik dalam mengklasifikasikan data dengan tingkat Akurasi sebesar 94,44%, Precision 97,30%, Recall 96,00%, dan F1-Score 96,64% untuk kelas *Buzzer*.

PENDANAAN

Penulis menyatakan bahwa tidak ada dukungan finansial yang diterima untuk pelaksanaan penelitian ini dan/atau penerbitan artikel ini.

KONFLIK KEPENTINGAN

Penulis menyatakan bahwa penelitian ini dilakukan tanpa adanya hubungan komersial atau finansial yang dapat ditafsirkan sebagai potensi konflik kepentingan.

KONTRIBUSI PENULIS

Esti Rahayu Putri: Conceptualization, methodology, data collection, data preprocessing, system development, data analysis, visualization, and writing the original draft. Wina Witanti: Supervision, methodology validation, critical review, interpretation of results, and manuscript revision. Edvin Ramadhan: Software development, implementation of the classification model, data analysis, evaluation, and manuscript revision. All authors contributed to reviewing and approving the final version of the manuscript.

PERNYATAAN PENGGUNAAN KECERDASAN ARTIFISIAL GENERATIF

Penulis menyatakan bahwa kecerdasan artifisial (AI) generatif digunakan dalam penyusunan manuskrip ini. Penulis mengakui penggunaan AI generatif untuk membantu meningkatkan kejelasan bahasa dan memperbaiki struktur tata bahasa manuskrip. Penggunaan tersebut dilakukan dengan mematuhi pedoman etika yang berlaku mengenai penggunaan dan pengungkapan AI generatif dalam penelitian ilmiah. Setiap penulis telah memberikan kontribusi intelektual yang substansial terhadap karya ini. Seluruh isi manuskrip telah ditelaah secara cermat untuk memastikan keakuratannya, dan penulis bertanggung jawab sepenuhnya atas integritas serta isi akhir manuskrip.

REFERENSI

- Akhtar, M. M., Bhuiyan, N. S., Masood, R., Ikram, M., & Kanhere, S. S. (2025). BotSSCL: Social Bot Detection with Self-Supervised Contrastive Learning. *Online Social Networks and Media*, 48, 100318. <https://doi.org/10.1016/j.osnem.2025.100318>
- Amalia, N. H., & Riki, R. (2024). *Implementasi Data Mining Pemasaran Produk Obat Menggunakan Algoritma Apriori Berbasis Web pada PT Murni Tritunggal Kendari*.
- Duan, J., Wu, S., Li, W., Bai, Q., Nguyen, M., & Jiang, J. (2026). BotDMM: Dual-channel multi-modal learning for LLM-driven bot detection on social media. *Information Fusion*, 127, 103758. <https://doi.org/10.1016/j.inffus.2025.103758>

- Ellaky, Z., & Benabbou, F. (2024). Political social media bot detection: Unveiling cutting-edge feature selection and engineering strategies in machine learning model development. *Scientific African*, 25, e02269. <https://doi.org/10.1016/j.sciaf.2024.e02269>
- Guyan, Q., Liu, Y., Liu, J., & Zhang, P. (2025). PEGNN: Peripheral-Enhanced graph neural network for social bot detection. *Expert Systems with Applications*, 278, 127294. <https://doi.org/10.1016/j.eswa.2025.127294>
- Hariyanto, E., Borman, R. I., & Rossi, F. (2024). *Visualizing Data Mining Classification for Machine Learning Education*.
- Huang, D., Song, J., & Zhang, X. (2025). Semi-Supervised Social Bot Detection with Relational Graph Attention Transformers and Characteristics of the social environment. *Information Fusion*, 118, 102956. <https://doi.org/10.1016/j.inffus.2025.102956>
- Latah, M. (2020). Detection of malicious social bots: A survey and a refined taxonomy. *Expert Systems with Applications*, 151, 113383. <https://doi.org/10.1016/j.eswa.2020.113383>
- Li, D., Shen, L., Guo, Q., Zhang, C., Li, J., Jiang, W., & Yu, M. (2025). BotLGT: Social bot detection based on LLM and graph transformer. *Neurocomputing*, 658, 131453. <https://doi.org/10.1016/j.neucom.2025.131453>
- Lingam, G., & Das, S. K. (2025). Social bot detection using variational generative adversarial networks with hidden Markov models in Twitter network. *Knowledge-Based Systems*, 311, 113019. <https://doi.org/10.1016/j.knosys.2025.113019>
- Liu, F., & Ma, R. (2025). From the past to the present: A social bot detection method based on spatio-temporal interactive perception. *Knowledge-Based Systems*, 322, 113712. <https://doi.org/10.1016/j.knosys.2025.113712>
- Lukin, E., Roberts, J. C., Berdik, D., Mugar, E., & Juola, P. (2023). *Adjectives and Adverbs as Stylometric Analysis Parameters*.
- Modupe, A., Celik, T., Marivate, V., & Olugbara, O. (2022). *Post-Authorship Attribution Using Regularized Deep Neural Network*.
- Nguyen, H.-D., Nguyen, D. Q., Nguyen, C.-D., To, P. T., Nguyen, D. H., Nguyen-Gia, H., Tran, L. H., Tran, A. Q., Dang-Hieu, A., Nguyen-Duc, A., & Quan, T. (2024). Supervised learning models for social bot detection: Literature review and benchmark. *Expert Systems with Applications*, 238, 122217. <https://doi.org/10.1016/j.eswa.2023.122217>
- Orabi, M., Mouheb, D., Al Aghbari, Z., & Kamel, I. (2020). Detection of Bots in Social Media: A Systematic Review. *Information Processing & Management*, 57(4), 102250. <https://doi.org/10.1016/j.ipm.2020.102250>
- Pal, A., Pranav, & Pradhan, M. (2023). Survey of fake news detection using machine intelligence approach. *Data & Knowledge Engineering*, 144, 102118. <https://doi.org/10.1016/j.datak.2022.102118>
- Raja, M. S., & Raj, L. A. (2022). Fake news detection on social networks using Machine learning techniques. *Materials Today: Proceedings*, 62(7), 4821–4827. <https://doi.org/10.1016/j.matpr.2022.03.351>
- Rini, D. S., & Prawiranegara, B. (2024). *Implementation of the Apriori Algorithm on Sales Data to Optimize Stock and Product Layout at Toko Senyum Kalsel*.
- Su, J., & Jiang, M. (2025). *Application and Optimization of Machine Learning and Data Mining Algorithms in the Detection of Financial Fraud*.
- Wardana, A., Syahroni, A., & Purwinarko, A. (2025). *Analysis of Market Basket Mining Using Apriori Algorithm on Transaction Data to Increase Sales*.