DETEKSI AKUN PALSU DI TWITTER BERBASIS GAYA BAHASA (STILOMETRI) MENGGUNAKAN XGBOOST

Christy, Hafiz Irsyad
Informatika, Universitas Multi Data Palembang
Jl. Rajawali No.14, 9 Ilir, Kec. Ilir Tim. II, Kota Palembang, Sumatera Selatan 30113, Indonesia
hafizirsyad@mdp.ac.id

ABSTRAK

Dengan meningkatnya penggunaan media sosial membuat Twitter menjadi ruang diskusi yang aktif, tetapi juga diiringi munculnya akun palsu yang menyebarkan hoaks dan memanipulasi opini publik. Permasalahan tersebut mendorong perlunya metode yang efektif untuk mengidentifikasi akun palsu secara otomatis. Penelitian ini bertujuan mengembangkan model deteksi akun palsu berbasis stilometri dengan memanfaatkan algoritma XGBoost. Data dikumpulkan melalui *web scraping tweet* berbahasa Indonesia, kemudian digabungkan per akun agar ciri linguistik lebih representatif. Label awal dibentuk menggunakan K-Means, dilanjutkan validasi oleh ahli bahasa Indonesia. Fitur yang digunakan meliputi stilometri, pola tanda baca, *char n-gram,* dan TF-IDF. Dataset dibagi menjadi 80% data latih dan 20% data uji dan dievaluasi menggunakan *confusion matrix*. Hasil pengujian menunjukkan bahwa penambahan fitur stilometri menunjukkan kecenderungan peningkatan model dalam membedakan akun asli dan palsu, dengan XGBoost menghasilkan nilai *F1-score* tertinggi, sebesar 0.982 pada uji skenario. Penelitian ini menunjukkan bahwa ciri linguistik tetap dapat dimanfaatkan secara efektif pada teks pendek seperti *tweet*, dan dapat menjadi pendekatan yang tepat untuk menjaga integritas percakapan publik di media sosial.

Kata kunci : *Deteksi Akun Palsu; Stilometri; XGBoost*

1. PENDAHULUAN

Perkembangan internet sebagai media komunikasi juga melatarbelakangi terbentuknya media sosial. Berdasarkan laporan We Are Social dalam Digital 2025, jumlah pengguna aktif media sosial di Indonesia diperkirakan mencapai lebih dari 181 juta orang [1]. Twitter (X) sering digunakan sebagai media diskusi isu-isu penting, tetapi kondisi ini juga sering dimanfaatkan oleh akun palsu untuk menyebarkan hoaks dan memanipulasi opini, yang berdampak pada kemurnian percakapan publik, keamanan siber, kesehatan mental, dan kualitas data sosial [2], [3], [4].

Permasalahan utama dalam deteksi akun palsu adalah tidak semua akun menunjukkan pola bahasa yang konsisten, terutama pada *platform* seperti Twitter yang didominasi teks pendek, bahasa informal, dan penggunaan slang. Berbagai penelitian terdahulu berfokus pada atribut metadata seperti profil pengguna, pola aktivitas, dan hubungan jaringan. Misalnya, sebuah penelitian menggunakan metadata serta model K-Means, Random Forest, AdaBoost, dan KNN, dengan hasil terbaik pada Random Forest dengan accuracy 98,55% [5]. Penelitian lain menggunakan algoritma Decision Tree dengan akurasi 99% [6]. Selain itu, penelitian yang memanfaatkan stilometri dan *semantic embedding* RoBERTa mampu mendeteksi perubahan penulis dengan akurasi hingga 99,2% [7].

Meskipun mampu menghasilkan akurasi tinggi, pendekatan berbasis metadata memiliki keterbatasan karena informasi akun relatif mudah dimanipulasi [8]. Oleh karena itu, diperlukan pendekatan alternatif yang lebih sulit dipalsukan, salah satunya melalui analisis gaya bahasa. Setiap orang memiliki karakteristik linguistik yang relatif unik, sehingga pendekatan stilometri berpotensi digunakan untuk membedakan akun asli dan akun palsu [9], termasuk pada tweet berbahasa Indonesia [10].

Seiring berkembangnya *machine learning*, algoritma seperti Random Forest, AdaBoost, KNN, Decision Tree, hingga XGBoost banyak digunakan. Penelitian ini bertujuan untuk mengembangkan model deteksi akun palsu di Twitter berbasis stilometri dengan menggunakan algoritma XGBoost. XGBoost dipilih karena mampu menangani data berukuran besar dan fitur numerik hasil ekstraksi stilometri secara optimal [11].

Meskipun demikian, penerapan stilometri pada Twitter masih memiliki tantangan karena karakteristik teks yang pendek, banyaknya penggunaan slang, dan variasi bahasa informal. Penelitian ini membatasi data pada *tweet* berbahasa Indonesia, mengekstraksi fitur stilometri yang relevan, serta menerapkan algoritma XGBoost untuk melakukan klasifikasi akun. Pendekatan ini diharapkan mampu meningkatkan akurasi deteksi akun palsu meskipun karakteristik teks Twitter bersifat singkat dan informal.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Sejumlah penelitian telah dilakukan untuk mendeteksi akun palsu di media sosial dengan berbagai pendekatan dan algoritma *machine learning*. Pendekatan awal umumnya menggunakan fitur metadata akun, seperti jumlah pengikut, aktivitas unggahan, dan interaksi pengguna. Beberapa penelitian menunjukkan bahwa algoritma Random

Forest dan Decision Tree mampu menghasilkan akurasi tinggi dalam membedakan akun asli dan akun palsu, bahkan mencapai lebih dari 98% pada dataset Twitter [5], [6].

Seiring berkembangnya penelitian, pendekatan berbasis linguistik mulai digunakan untuk melengkapi keterbatasan metadata yang relatif mudah dimanipulasi. Stilometri dimanfaatkan untuk menganalisis pola gaya bahasa penulis berdasarkan karakteristik linguistik yang cenderung konsisten. Penelitian sebelumnya menunjukkan bahwa fitur stilometri efektif digunakan untuk mendeteksi perbedaan penulis serta mengidentifikasi pola bahasa pada teks pendek, termasuk pada konten media sosial berbahasa Indonesia [7], [10].

Selain itu, beberapa studi mengenai hoaks di media sosial mengungkapkan bahwa penggunaan tanda baca berlebihan, huruf kapital, serta ekspresi emosional ekstrem merupakan ciri linguistik yang sering muncul pada konten palsu [12]. Pola tersebut menjadi dasar penting dalam ekstraksi fitur stilometri. Algoritma XGBoost juga mulai banyak digunakan karena kemampuannya mengolah fitur numerik berdimensi tinggi dan meningkatkan akurasi klasifikasi melalui mekanisme *boosting* dan regularisasi [13], [14]. Oleh karena itu, penelitian ini mengombinasikan pendekatan stilometri dan algoritma XGBoost untuk mendeteksi akun palsu di Twitter.

2.2. Machine Learning

Machine learning merupakan cabang kecerdasan buatan yang memungkinkan sistem belajar dari data untuk membuat prediksi atau keputusan baru [15]. Metode ini dapat bekerja pada data berlabel (*supervised learning*) maupun tanpa label (*unsupervised learning*), serta banyak digunakan pada pengenalan wajah, analisis media sosial, dan klasifikasi teks.

2.3. Akun Palsu

Akun palsu adalah akun yang menggunakan identitas tidak asli dan biasanya digunakan untuk penipuan, manipulasi opini publik, atau penyebaran konten berbahaya. Akun palsu dapat berupa bot, akun manusia palsu (*troll*), maupun *cyborg* (kombinasi manusia dan bot) [16].

2.4. Stilometri

Stilometri merupakan metode kuantitatif untuk menganalisis gaya penulisan berdasarkan pola linguistik seperti frekuensi kata, panjang kalimat, dan penggunaan tanda baca. Stilometri digunakan sebagai dasar ekstraksi fitur. Penggunaan Stilometri dapat dilihat pada tabel 1 dibawah ini.

Tabel 1. Penggunaan jenis tanda baca

No.	Tanda Baca	Makna / Nada Emosional
1.	Tanda Seru (!)	Menunjukkan emosi yang meningkat seperti kegembiraan, urgensi, kemarahan, atau antusiasme. Penggunaan yang berlebihan dapat memberi kesan agresif atau dramatis.
2.	Tanda Tanya (?)	Menandakan pertanyaan yang mendorong respon. Penggunaan berlebihan dapat menunjukkan rasa ingin tahu, keraguan, atau konfrontasi.
3.	Elipsis (…)	Mengindikasikan keraguan, ketidakpastian, atau percakapan yang dibiarkan terbuka. Penggunaan berlebihan dapat memberi kesan misterius atau pasif-agresif.
4.	Koma (,)	Menjaga aliran kalimat, memberikan jeda kecil dan membuat teks mudah dibaca tanpa memberikan emosi yang kuat.
5.	Titik (.)	Memberikan finalitas, kesimpulan, atau nada netral. Dalam konteks informal media sosial, titik di akhir kalimat kadang dianggap dingin atau tegas.

Berdasarkan Tabel 1, tanda baca tidak hanya berfungsi secara struktural dalam teks, tetapi juga berperan dalam menyampaikan nada, emosi, dan intensi penulis [12]. Pada Tabel 2 dapat dilihat bahwasan tanda baca yang memiliki indikator hoaks.

Tabel 2. Tanda baca indikator hoaks

No.	Pola Tanda Baca	Makna dalam Hoaks	Contoh dari Studi
1.	Beberapa tanda seru berurutan (!!!)	Menekankan urgensi, provokasi, kemarahan, atau mencoba membangkitkan kepanikan	Pliiis !!!! jangan katakan Kalimantan ndeso…
2.	Tanda tanya berlebihan (???)	Menyiratkan kecurigaan, memancing rasa ingin tahu berlebihan, dan membangun narasi konspiratif	*Boleh dicek di sini???*
3.	Kombinasi tanda seru dan tanya (!?, ?!)	Menunjukkan serangan emosional, bias, dan nada marah/menuduh	JOKOWI TUNJUK RISMA GANTIKAN ANIES !?.;
4.	Penggunaan huruf besar di seluruh judul (ALL CAPS) dikombinasikan dengan tanda baca berlebihan	Meningkatkan intensitas dramatis	SEMAKIN MEM4NAS !!!

Berdasarkan Tabel 2, salah satu ciri linguistik utama pada konten hoaks di media sosial adalah penggunaan tanda baca secara berlebihan. Pola ini muncul karena pembuat hoaks berupaya menarik perhatian, membangkitkan emosi pembaca, serta memicu penyebaran informasi secara cepat [17].

2.5. TF-IDF

Term Frequency–Inverse Document Frequency (TF-IDF) adalah metode pemberian bobot kata berdasarkan frekuensi kemunculannya dalam suatu dokumen dan kelangkaannya pada seluruh dokumen [18]. TF menggambarkan seberapa sering kata muncul dalam satu dokumen, sedangkan IDF menunjukkan tingkat keberbedaan kata tersebut pada kumpulan dokumen. TF-IDF digunakan untuk mengubah teks menjadi fitur numerik yang dapat diproses oleh algoritma *machine learning*.

$$tf(t,d) = \frac{f(t,d)}{\Sigma f(d)} \quad (1)$$

dimana:
t = *term* (kata)
d = dokumen
$f(t,d)$ = jumlah kemunculan *term* t dalam dokumen d
$\Sigma f(d)$ = jumlah seluruh kata dalam dokumen d

$$idf(t) = log\left(\frac{N}{df(t)}\right) \quad (2)$$

dimana:
N = jumlah total dokumen
$df(t)$ = jumlah dokumen yang mengandung kata t

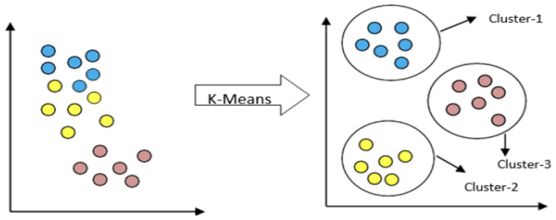
$$tf - idf(t,d) = tf(t,d) * idf(t) \quad (3)$$

2.6. K-Means

K-Means merupakan algoritma *clustering* yang membagi data ke dalam k kelompok berdasarkan kedekatan jaraknya terhadap *centroid*. Algoritma bekerja dengan menentukan pusat awal secara acak, menghitung jarak setiap data ke setiap *centroid*, dan memperbarui *centroid* hingga pembagian klaster stabil [19]. Setiap data akan dihitung jaraknya ke setiap pusat dengan rumus Euclidean yang dapat dilihat pada Persamaan (4).

$$D = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2} \quad (4)$$

dimana:
x_1 dan x_2 = koordinat titik data
c_1 dan c_2 = koordinat *centroid*



Gambar 1. K-Means *clustering*

Posisi pusat klaster diperbarui berdasarkan rata-rata data, dan proses pengelompokan diulang hingga hasil klasterisasi stabil sehingga data dalam satu klaster memiliki kemiripan tinggi dan berbeda dari klaster lainnya yang dapat dilihat pada Gambar 1.

2.7. XGBoost

XGBoost (*Extreme Gradient Boosting*) adalah algoritma *boosting* berbasis pohon keputusan yang meningkatkan akurasi dengan membangun banyak pohon secara bertahap. Setiap pohon memperbaiki kesalahan pohon sebelumnya menggunakan optimasi *gradient* dan regularisasi sehingga mengurangi risiko *overfitting* serta efektif pada dataset berukuran besar [13], [14]. Fungsi objektif pada XGBoost yang digunakan dijelaskan pada Persamaan (5).

$$Obj = L(y_i, \hat{y}_i) + \lambda\Sigma|w| + \frac{1}{2}\gamma\Sigma w^2 \quad (5)$$

Keterangan:
w = bobot yang mewakili kekuatan suatu fitur terhadap hasil prediksi
λ = koefisien Regularisasi L1
γ = koefisien Regularisasi L2
y_i = label sebenarnya dari data
$\hat{y}_i$ = nilai hasil prediksi dari model XGBoost untuk data ke-i

2.8. Confusion Matrix

Confusion matrix adalah tabel evaluasi model klasifikasi yang menampilkan jumlah prediksi benar dan salah, terdiri dari *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dari tabel ini dapat dihitung metrik performa seperti *accuracy*, *precision*, *recall*, dan *F1-score* [20], [21]. Keempat elemen dalam *confusion matrix* pada Tabel 3.

Tabel 3. *Confusion matrix*

		Prediction	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Keterangan:
- TP (*True Positive*) = data bernilai positif yang diprediksi benar sebagai positif
- TN (*True Negative*) = data bernilai negatif yang diprediksi benar sebagai negatif
- FP (*False Positive*) = data bernilai negatif yang diprediksi salah sebagai positif
- FN (*False Negative*) = data bernilai positif yang diprediksi salah sebagai negatif

Berdasarkan nilai-nilai tersebut, metrik seperti *accuracy*, *precision*, *recall*, dan F1-*score* dapat dihitung dengan rumus berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

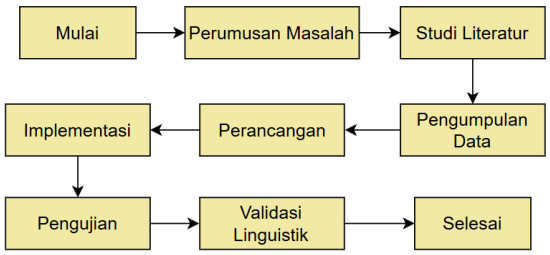
$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

3. METODE PENELITIAN

Tahapan metodologi penelitian dapat dilihat pada Gambar 2.



Gambar 2. Tahapan metodologi penelitian

3.1. Perumusan Masalah

Tahapan ini diawali dengan identifikasi masalah meningkatnya penyebaran informasi palsu di media sosial yang berasal dari akun palsu. Permasalahan difokuskan pada bagaimana algoritma XGBoost dapat diterapkan untuk membedakan akun asli dan palsu berdasarkan ciri linguistik yang muncul dalam teks *tweet*.

3.2. Studi Literatur

Studi literatur dilakukan dengan mengkaji jurnal dan buku terkait analisis stilometri, klasifikasi teks, serta algoritma *machine learning* seperti K-Means dan XGBoost. Pada tahap ini juga dipelajari konsep TF-IDF, ekstraksi fitur, dan metrik evaluasi seperti *accuracy, precision, recall,* dan *F1-score.*

3.3. Pengumpulan Data

Dataset dikumpulkan melalui metode *web scraping* menggunakan *tweet-harvest* melalui Node.js dengan kata kunci "demo", berbahasa Indonesia, dan rentang waktu 3–4 bulan pada periode 2022–2025. Pemilihan kata kunci "demo" dipilih karena isu demonstrasi sering melibatkan akun palsu dan penyebaran opini terkoordinasi, meskipun berpotensi menimbulkan bias topik dimana model dapat lebih banyak mempelajari pola linguistik spesifik terkait isu demonstrasi daripada pola umum akun palsu. *Scraping* dilakukan empat (4) kali karena keterbatasan API dan disimpan dalam file CSV yang kemudian digabungkan menjadi satu (1) file *TwitterDemo_Combined_AllPeriods.csv*. Selanjutnya, dilakukan *data cleaning* dengan menghapus duplikasi berdasarkan kolom *full_text*.

Setelah penggabungan dan pembersihan data, diperoleh 1.132 *tweet* dengan 15 atribut. Dataset yang digunakan merupakan data publik, dengan identitas pengguna dianonimkan untuk menjaga privasi. Data hanya digunakan untuk kepentingan penelitian ilmiah. Validasi akun palsu dilakukan oleh dosen Bahasa Indonesia berdasarkan indikator seperti kelengkapan profil, riwayat akun dan aktivitas, rasio jumlah pengikut, frekuensi unggahan, serta indikator stilometri lainnya.

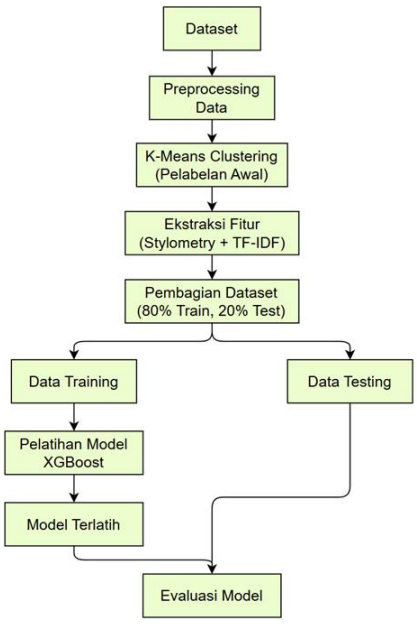
3.4. Perancangan

Data kemudian dikelompokkan berdasarkan akun pengguna (*user_id_str*) agar setiap akun hanya memiliki satu representasi teks berupa gabungan semua *tweet*-nya agar fitur stilometri tetap representatif. Hal ini dilakukan untuk mengekstraksi gaya bahasa dari setiap akun, bukan dari setiap *tweet* tunggal.

Meskipun *tweet* bersifat pendek dan informal, fitur stilometri tetap dapat direpresentasikan secara akurat dengan menggabungkan seluruh *tweet* per akun dan memfokuskan ekstraksi pada fitur untuk teks pendek seperti pola tanda baca, kapitalisasi, emoji, pemanjangan huruf, serta *char n-gram.*

Karena data tidak memiliki label, digunakan K-Means dengan jumlah klaster k = 2, yang merepresentasikan dua kelas, yaitu akun asli dan palsu. Pendekatan TF-IDF *char n-gram* dipilih karena mampu menangani teks campuran bahasa dan kesalahan penulisan serta lebih robust terhadap variasi ejaan [22], mampu menangkap pola linguistik informal [23], dan tidak bergantung pada kosakata baku [24] .

Setelah itu, dilakukan ekstraksi fitur stilometri dan TF-IDF untuk menghasilkan representasi numerik dari teks untuk keperluan deteksi. Semua fitur ini akan digabungkan menjadi satu matriks fitur (X) dengan label hasil klasterisasi (y) yang kemudian dibagi menjadi 80% data latih dan 20% data uji menggunakan *train_test_split* dengan *stratify* untuk menjaga distribusi label. Selama proses pelatihan, model XGBoost akan belajar memetakan kombinasi fitur numerik dari teks (X) ke label biner (y) agar bisa memprediksi apakah sebuah akun termasuk asli atau palsu.



Gambar 3. Skema perancangan

Evaluasi dilakukan menggunakan *confusion matrix* dan beberapa metrik kinerja, yaitu *accuracy, precision, recall,* dan F1-*score*. PCA (2D) diterapkan

untuk mereduksi dimensi TF-IDF dan menampilkan distribusi klaster. Analisis *scatter plot* digunakan untuk memeriksa pemisahan akun asli dan palsu, sehingga validitas klaster dapat dievaluasi secara visual. Visualisasi ini akan menilai apakah terjadi tumpang tindih yang besar (indikasi klaster kurang baik). Selain itu, dilakukan analisis *feature importance* untuk mengetahui fitur mana yang paling berpengaruh terhadap hasil prediksi XGBoost. Skema perancangan dapat dilihat pada Gambar 3.

3.5. Implementasi

Model diimplementasikan menggunakan antarmuka sederhana berbasis *ipywidgets* di Google Colab. Pengguna dapat memasukkan teks tweet dan sistem menampilkan hasil prediksi akun asli atau palsu beserta nilai probabilitasnya.

3.6. Pengujian

Pengujian dilakukan menggunakan data uji sebesar 20% dari dataset untuk mengukur kinerja model. Evaluasi menggunakan *confusion matrix, accuracy, precision, recall,* dan *F1-score.*

Pengujian dilakukan pada dua skenario, yaitu XGBoost dengan fitur TF-IDF dan stilometri, serta XGBoost dengan TF-IDF saja. Fokus evaluasi pada *F1-score* untuk membuktikan bahwa penambahan fitur stilometri meningkatkan kemampuan deteksi akun palsu.

3.7. Validasi Linguistik

Untuk menghasilkan pelabelan yang bias dari hasil klaster yang diperoleh K-Means, maka membutuhkan validasi dari dosen ahli Bahasa Indonesia. Validasi dilakukan dengan menganalisis karakteristik linguistik dan perilaku dari akun seperti pada tabel 1 dan 2.

4. HASIL DAN PEMBAHASAN

Setelah dilakukan pembersihan data duplikasi berdasarkan *full_text*, jumlah *tweet* yang tersisa adalah 1.132 *tweet* dari berbagai akun. Selanjutnya, *tweet* dikelompokkan berdasarkan *user_id_str* sehingga setiap akun hanya memiliki satu dokumen teks berupa gabungan seluruh *tweet* miliknya. Langkah ini bertujuan menjaga konsistensi gaya penulisan agar fitur stilometri yang diekstraksi benar-benar mewakili karakter linguistik tiap akun. Hasil pengelompokan menghasilkan 1.056 akun unik, yang kemudian digunakan sebagai objek klasifikasi.

Tabel 4. Atribut dataset

Atribut (Kolom)	Jenis Data	Keterangan
conversation_id_str	int64	ID unik untuk seluruh percakapan. Semua *tweet* dalam satu percakapan memiliki *conversation_id_str* yang sama.
created_at	object	Waktu dan tanggal *tweet* dibuat.
favorite_count	int64	Jumlah suka (*like*) yang diterima oleh *tweet* tersebut.
full_text	object	Isi lengkap teks dari *tweet* yang menjadi kolom utama untuk analisis teks.
id_str	int64	ID unik dari *tweet* itu sendiri dalam bentuk teks.
image_url	object	URL gambar (jika *tweet* mengandung gambar) dan bisa kosong jika tidak ada gambar.
in_reply_to_screen_name	object	*Username* akun yang dibalas oleh *tweet* tersebut. Jika *tweet* bukan balasan, biasanya nilainya NaN atau *None*.
lang	object	Bahasa yang terdeteksi dalam *tweet*. Contoh: en (Inggris), id (Indonesia).
location	float64	Lokasi pengguna (diambil dari profil pengguna, bukan lokasi GPS *tweet*).
quote_count	int64	Berapa kali jumlah *tweet* tersebut dikutip (*quote tweet*) oleh pengguna lain.
reply_count	int64	Jumlah balasan (*reply*) yang diterima oleh *tweet* tersebut.
retweet_count	int64	Jumlah *retweet* terhadap *tweet* tersebut.
tweet_url	object	Tautan langsung (URL) ke *tweet* di Twitter.
user_id_str	int64	ID unik pengguna (user) yang membuat *tweet* tersebut.
username	float64	Nama pengguna dari akun Twitter yang membuat *tweet*.

Berdasarkan Tabel 4, setiap akun memiliki 15 atribut awal, namun hanya atribut *full_text* yang diproses untuk ekstraksi fitur TF-IDF dan stilometri.

Kemudian, dilakukan proses pelabelan menggunakan K-Means dengan jumlah klaster *k = 2* (akun asli dan akun palsu). Dari proses *clustering*, diperoleh hasil label 0 (akun asli) sebanyak 781 dan label 1 (akun palsu) sebanyak 275 data. hasil *clustering* ini tidak digunakan sebagai label akhir, melainkan sebagai *pseudo-label* awal dalam skema *semi-supervised learning*. Penentuan klaster sebagai akun palsu atau asli dilakukan melalui validasi linguistik oleh ahli Bahasa Indonesia. Dari hasil validasi oleh ahli Bahasa Indonesia menghasilkan cluster 1 merupakan akun palsu ditentukan berdasarkan analisis stilometri rata-rata setiap klaster. Cluster 1 menunjukkan nilai lebih tinggi pada kategori penggunaan tanda seru berlebihan, tanda tanya berurutan, huruf kapital tidak wajar, dan elipsis berulang. Pola ini konsisten dengan temuan linguistik pada konten provokatif atau manipulatif sebagaimana dijelaskan pada Tabel 2.

Ekstraksi fitur stilometri yang dilakukan sebanyak 21 fitur, berupa: fitur leksikal (jumlah kata, rata-rata panjang kata, *type token ratio (TTR)*, kapitalisasi), fitur tanda baca (urutan tanda seru (!!!),

tanda tanya (???), tanda baca campuran (!?, ?!), elipsis, pengulangan huruf), fitur media sosial (jumlah *hashtag*, *mention*, link, emoji), fitur POS Tagging (proporsi *noun*, *verb*, *adjective*, *adverb*, *pronoun*), dan *sentiment* menggunakan *TextBlob polarity*. Dataset kemudian dibagi menjadi 80% data latih dan 20% data uji dan menghasilkan pembagian data yang dapat dilihat pada Tabel 5.

Tabel 5. Hasil pembagian dataset

Kelas	Data Latih	Data Uji
Label 0 (asli)	624	157
Label 1 (palsu)	220	55

Selanjutnya kedua label ditinjau ulang oleh validator ahli bahasa untuk memastikan bahwa klaster tidak terbalik dan sesuai karakteristik akun palsu pada bahasa Indonesia. Berdasarkan pemeriksaan, mayoritas akun dalam klaster dengan ciri penggunaan tanda baca berlebihan, huruf kapital masif, dan frekuensi *tweet* tidak wajar ditetapkan sebagai label akun palsu.

Karena dataset memperlihatkan ketidakseimbangan antara kelas asli dan palsu, ,maka digunakan parameter *scale_pos_weight* = 2.84 pada XGBoost untuk menyeimbangkan bobot kelas minoritas. Model XGBoost dilatih menggunakan gabungan fitur stilometri dan TF-IDF tereduksi (SVD). Parameter *max_depth* = 5 digunakan untuk membatasi kedalaman pohon agar model tidak terlalu kompleks dalam mempelajari pola yang sebenarnya tidak signifikan. Selanjutnya, *subsample* = 0.7 dan *colsample_bytree* = 0.7 digunakan untuk membatasi sebagian data dan sebagian fitur yang digunakan setiap pohon, sehingga model tidak terlalu bergantung pada sampel atau fitur tertentu. Komponen regularisasi juga sangat berpengaruh, yaitu *reg_alpha* = 2.0 (L1) dan *reg_lambda* = 5.0 (L2), yang berfungsi menekan bobot fitur yang terlalu besar dan mengurangi kompleksitas model secara keseluruhan. Selain itu, *learning_rate* = 0.07 menjaga agar setiap pohon hanya memberikan kontribusi kecil sehingga proses pembelajaran berlangsung lebih stabil dan tidak terlalu agresif. Kombinasi parameter-parameter ini akan membantu XGBoost tetap kuat dalam mempelajari pola tanpa *overfitting*.

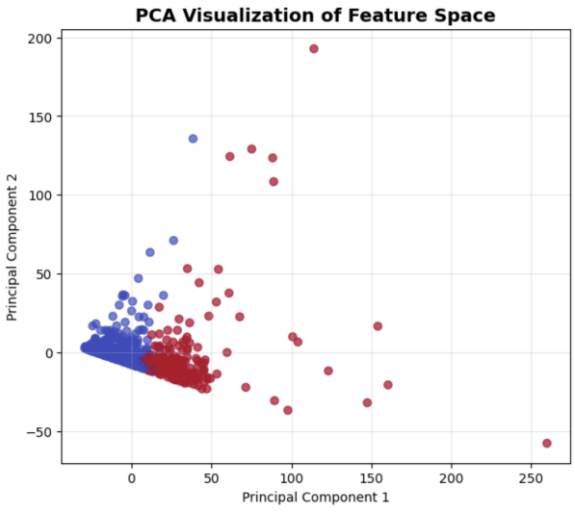
Tabel 6. Hasil evaluasi model

Model	*Accuracy*	*Precision*	*Recall*	*F1-score*
XGBoost (TF-IDF + Stilometri)	0.990566	0.964912	1.000000	0.982143
XGBoost (TF-IDF)	0.867925	0.728814	0.781818	0.754386

Berdasarkan Tabel 6, model XGBoost menjadi model dengan kinerja paling tinggi dengan *accuracy* sebesar 0.990566 (≈99%) yang berarti dari seluruh data uji, sekitar 99% prediksi yang dibuat sudah benar. Nilai *precision* sebesar 0.964912 (≈96%) menunjukkan bahwa hampir semua *tweet* yang diprediksi sebagai palsu memang benar-benar palsu, sehingga kesalahan menandai *tweet* normal sebagai palsu sangat kecil. Sementara itu, *recall* sebesar 1.0 (100%) menandakan bahwa XGBoost mampu menemukan seluruh *tweet* palsu tanpa ada yang terlewat. Kombinasi *precision* dan *recall* yang tinggi menghasilkan *F1-score* sebesar 0.982143, yang menandakan performa model ini sangat seimbang.

Berdasarkan uji skenario, terlihat penurunan performa akurasi XGBoost sebesar 12% (dari 0.99 menjadi 0.867) jika hanya menggunakan fitur TF-IDF saja. Hal ini membuktikan bahwa penggunaan TF-IDF untuk pendekatan berbasis representasi teks belum menjamin hasil prediksi yang akurat. Temuan ini juga menunjukkan bahwa fitur stilometri yang menangkap pola dan gaya penulisan khas pengguna, sangat penting untuk membedakan akun asli dan palsu.

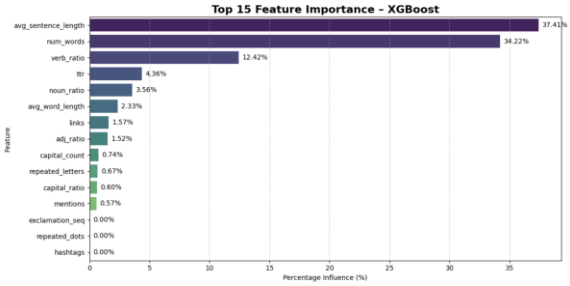
Secara keseluruhan, hasil ini menunjukkan bahwa model XGBoost dengan fitur TF-IDF dan stilometri memberikan performa terbaik. Sementara penggunaan TF-IDF saja tidak cukup kuat untuk menghasilkan prediksi yang akurat, sehingga fitur stilometri berperan penting dalam meningkatkan kemampuan deteksi spam.



Gambar 4. Visualisasi PCA

Grafik di atas menampilkan hasil visualisasi data ke dalam dua komponen utama menggunakan PCA untuk melihat pola penyebaran data dalam bentuk yang lebih sederhana, sehingga perbedaan karakteristik antar kelompok bisa terlihat lebih jelas. Pada grafik, titik biru mewakili akun asli, sedangkan titik merah mewakili akun palsu. Terlihat bahwa akun asli (biru) cenderung berkumpul lebih rapat di area kiri bawah grafik. Pola ini menunjukkan bahwa fitur-fitur yang dimiliki akun asli memiliki kemiripan yang konsisten, sehingga posisinya di ruang PCA tidak terlalu menyebar jauh yang berarti tidak menunjukkan variasi ekstrem. Sedangkan, akun palsu (merah) tampak menyebar lebih luas ke berbagai arah. Penyebaran yang lebar menunjukkan bahwa akun palsu memiliki ciri-ciri yang lebih bervariasi dan tidak

seragam. Beberapa titik bahkan berada cukup jauh dari pusat kumpulan data, menunjukkan bahwa keberadaan akun-akun palsu yang memiliki pola perilaku sangat berbeda dari akun pada umumnya. Meski masih ada area di mana keduanya saling tumpang tindih, pola umum grafik menunjukkan bahwa akun palsu memiliki rentang variasi fitur yang lebih luas, sedangkan akun asli lebih terkonsentrasi pada satu area tertentu.



Gambar 5. *Feature importance*

Visualisasi *feature importance* menunjukkan bahwa model sangat bergantung pada beberapa fitur utama untuk membedakan akun asli dan akun palsu. Fitur paling berpengaruh adalah *avg_sentence_length* dengan kontribusi 37.41%, diikuti num_words sebesar 34.22%. Kedua fitur ini menggambarkan panjang dan struktur tulisan yang berbeda antara akun asli dan palsu, dimana akun palsu sering menulis kalimat lebih panjang, tidak natural, atau terlalu pendek dan bertele-tele.

Fitur berikutnya adalah *verb_ratio* dengan pengaruh 12.42%, menunjukkan bahwa proporsi kata kerja juga menjadi sinyal penting. Setelah itu, fitur seperti *ttr* (4.36%), *noun_ratio* (3.56%), dan *avg_word_length* (2.33%) yang mencerminkan gaya penulisan alami atau buatan. Fitur lain seperti jumlah tautan, rasio kata sifat, serta pola penulisan (huruf kapital, huruf berulang, dan sebagainya) memberi pengaruh kecil namun tetap membantu model menangkap pola yang tidak wajar. Secara keseluruhan, hasil ini memperlihatkan bahwa XGBoost terutama mengandalkan pola dan struktur penulisan, khususnya panjang kalimat dan jumlah kata, untuk membedakan kedua jenis akun.

Temuan ini menunjukkan bahwa pendekatan berbasis gaya bahasa dapat menjawab keterbatasan yang diidentifikasi pada bagian pendahuluan yang banyak bergantung pada metadata akun dalam mendeteksi akun palsu di media sosial. Meskipun tidak menggunakan informasi metadata akun, seperti jumlah pengikut, usia akun, atau pola jaringan, model XGBoost tetap menghasilkan kinerja klasifikasi yang baik. Hal ini menunjukkan bahwa pola linguistik dalam gaya penulisan akun dapat menjadi pendekatan yang efektif untuk membedakan akun asli dan palsu bahkan pada teks singkat dan informal.

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa stilometri efektif untuk mendeteksi akun palsu di Twitter, terutama pada *tweet* berbahasa Indonesia. Pola linguistik seperti panjang kalimat, jumlah kata, dan variasi jenis kata terbukti mampu membedakan akun asli dan palsu meskipun teks bersifat pendek dan informal. Penerapan XGBoost dengan kombinasi fitur stilometri dan TF-IDF menunjukkan performa unggul dengan nilai *accuracy* 99% serta *precision* dan *recall* yang tinggi, menunjukkan bahwa XGBoost mampu memanfaatkan ciri linguistik secara optimal dalam proses klasifikasi. Proses *clustering* dengan K-Means dan validasi oleh ahli bahasa ikut memastikan bahwa pelabelan data sudah sesuai. Hasil pengujian juga menunjukkan bahwa penggunaan fitur TF-IDF saja belum cukup untuk menghasilkan performa deteksi yang optimal dengan nilai *accuracy* sebesar 86%, serta *precision* dan *recall* yang lebih kecil dibandingkan dengan kombinasi stilometri.

Meski demikian, penelitian ini masih memiliki keterbatasan pada ukuran dan variasi dataset, sehingga penelitian berikutnya diharapkan menggunakan dataset yang lebih besar dengan beragam topik dan bahasa agar cakupan model lebih luas. Selain itu, penerapan uji *cross-topic* dapat dilakukan untuk melihat apakah model tetap efektif dalam berbagai topik, penggunaan label manual parsial juga disarankan untuk meningkatkan kualitas paleblan awal dan validasi model, serta perbandingan dengan algoritma *machine learning* lainnya untuk melihat kinerja model XGBoost. Model juga perlu diuji pada platform media sosial berbeda untuk melihat konsistensinya.

DAFTAR PUSTAKA

[1] We Are Social Indonesia, “Digital 2025,” We Are Social. [Online]. Available: https://wearesocial.com/id/blog/2025/02/digital-2025/

[2] S. Sugihartono, “The Power of Social Media to Influence Political Views and Geopolitical Issues: TikTok, X and Instagram,” Modern Diplomacy. [Online]. Available: https://moderndiplomacy.eu/2024/09/19/the-power-of-social-media-to-influence-political-views-and-geopolitical-issues-tiktok-x-and-instagram/

[3] JPNN.com, “Twitter Hapus Ratusan Akun Palsu Pendukung Pemerintah.” [Online]. Available: https://www.jpnn.com/news/twitter-hapus-ratusan-akun-palsu-pendukung-pemerintah-indonesia

[4] TweetDelete, “Akun Twitter Palsu: Menjelajahi Fenomena.” [Online]. Available: https://tweetdelete.net/id/resources/fake-twitter-accounts-exploring-the-phenomenon/

[5] A. P. A. Zahra, R. Hendradi, and F. Effendy, “Deteksi Akun Palsu Media Sosial X (Twitter) Berbasis Machine Learning,” in Prosiding

Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika (SNESTIK), 2025, pp. 319–326. [Online]. Available: https://ejurnal.itats.ac.id/snestik/article/view/7343/4781

[6] R. Taufik, R. Jimah, and A. Solichin, "Implementasi dan Analisis Model Machine Learning Decision Tree untuk Deteksi Akun Palsu di Twitter," J. Media Inform. Budidarma, vol. 8, no. 2, pp. 797–809, 2024, [Online]. Available: https://pdfs.semanticscholar.org/6553/507ff46937f7e2e73a48a2a77e2e50df4926.pdf

[7] T. Kumarage, J. Garland, A. Bhattacharjee, K. Trapeznikov, S. Ruston, and H. Liu, "Stylometric Detection of AI-Generated Text in Twitter Timelines," 2023. [Online]. Available: https://arxiv.org/pdf/2303.03697

[8] A. F. Fauzi, "Identifikasi Fake-Account Twitter terhadap Social Engineering Pretexting pada Akun Bank dan E-Wallet di Indonesia Menggunakan Metode Naive Bayes, Neural-Network & SVM," Universitas Mercu Buana Jakarta, 2022. [Online]. Available: https://repository.mercubuana.ac.id/70965/

[9] A. I. Rahmawati, "Linguistic Features of Fake Accounts in Twitter: A Corpus Study," Universitas UIN Sunan Ampel Surabaya, 2020. [Online]. Available: http://digilib.uinsa.ac.id/43741/2/Annisa Ismi Rahmawati_A3216103.pdf#:~:text=akun-akun palsu sering menggunakan kata-kata non formal

[10] A. Fathin, Y. Sari, and M. O. Nurfajri, "Authorship Profiling Tweet Berbahasa Indonesia Berdasarkan Fitur Stylometry Menggunakan Support Vector Machine (SVM)," 2025. [Online]. Available: https://etd.repository.ugm.ac.id/penelitian/detail/258050#:~:text=menggabungkan fitur stylometry seperti jumlah karakter%2C rasio huruf kapital%2C jumlah tanda baca%2C emoji

[11] C. Valentino and L. A. A. R. Putri, "Analisis Kinerja XGBoost Menggunakan Bayesian Optimization dalam Prediksi Harga Ethereum," J. Nas. Teknol. Inf. dan Apl., vol. 3, no. 4, pp. 795–804, 2025, [Online]. Available: https://ejournal2.unud.ac.id/index.php/jnatia/article/view/16%0A

[12] A. Y. Tenriawali, Sumiaty, Hanapi, I. Hajar, and N. AR, "Hoax Language Style in Indonesian Social Media," Uniqbu J. Soc. Sci., vol. 2, no. 1, pp. 62–68, 2021, [Online]. Available: https://media.neliti.com/media/publications/545539-none-893f1203.pdf

[13] A. A. Saputra, B. N. Sari, and C. Rozikin, "Penerapan Algoritma Extreme Gradient Boosting (XGBoost) untuk Analisis Risiko Kredit," J. Ilm. Wahana Pendidik., vol. 10, no. 7, pp. 27–36, 2024, [Online]. Available: https://www.jurnal.peneliti.net/index.php/JIWP/article/view/6670

[14] J. M. A. S. Dachi and P. Sitompul, "Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit," J. Ris. Rumpun Mat. Dan Ilmu Pengetah. Alam, vol. 2, no. 2, pp. 87–103, 2023, [Online]. Available: https://prin.or.id/index.php/JURRIMIPA/article/view/1470

[15] P. D. Kusuma, Machine Learning Teori, Program, dan Studi Kasus. Deepublish, 2020. [Online]. Available: https://books.google.co.id/books?id=wDVaEQAAQBAJ

[16] S. Refalia and B. Prasetyo, "Pembuktian Akun Palsu terhadap Selebgram yang Diduga Melakukan Promosi Judi Online," J. Huk. Lex Gen., vol. 5, no. 7, pp. 1–14, 2024, [Online]. Available: https://ojs.rewangrencang.com/index.php/JHLG/article/view/713/317

[17] Y. Lubis, A. Prasetio, R. Maulana, and S. Rahmadina, "The Use of Punctuation in Writing Captions on Social Media," J. Pendidik. Tambusai, vol. 9, no. 1, pp. 2852–2861, 2025, [Online]. Available: https://jptam.org/index.php/jptam/article/download/24712/16833/41930

[18] L. Suryani and K. Edy, "Pengembangan Aplikasi 'Lost & Found' Berbasis Android dengan Menggunakan Metode Term Frequency–Inverse Document Frequency (TF-IDF) dan Cosine Similarity," Electro Luceat, vol. 6, no. 2, pp. 190–204, 2020, [Online]. Available: https://www.poltekstpaul.ac.id/jurnal/index.php/jelekn/article/view/232%0A

[19] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering dalam Evaluasi Hasil Pembelajaran Siswa," J. Ilm. Inform. dan Ilmu Komput., vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: https://e.publication.diskoplampung.com/index.php/jima-ilkom/article/view/26

[20] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier dan Confusion Matrix pada Analisis Sentimen Berbasis Teks pada Twitter," J. Sains Komput. Inform., vol. 5, no. 2, pp. 697–711, 2021, [Online]. Available: https://www.academia.edu/download/106431308/348.pdf

[21] E. Fauziningrum and E. I. Suryaningsih, "Evaluasi dan Prediksi Penguasaan Bahasa Inggris Maritim Menggunakan Metode Decision Tree dan Confusion Matrix (Studi Kasus di Universitas Maritim AMNI)," in Prosiding Kemaritiman 2021, Semarang: Universitas Maritim AMNI, 2021. [Online]. Available: http://repository.unimar-amni.ac.id/3837/

[22] A. Kulmizev et al., "The Power of Character N-grams in Native Language Identification," pp. 382–389, 2017, [Online]. Available: https://aclanthology.org/W17-5043/

[23] J. Piskorski and G. Jacquet, "TF-IDF Character N -grams versus Word Embedding-based Models for Fine-grained Event Classification : A Preliminary study," pp. 26–34, 2020, [Online]. Available: https://aclanthology.org/2020.aespen-1.6/

[24] J. Kruczek, P. Kruczek, and M. K. B, Are n-gram Categories Helpful in Text Classification ? Springer International Publishing, 2020. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7302864/