

**PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK
MENGELOMPOKKAN ANOMALI KOMENTAR NORMAL,
SPAM, DAN BOT PADA PLATFORM TIKTOK**



SKRIPSI

Diajukan Sebagai Syarat untuk Memperoleh Gelar Sarjana Komputer
(S.Kom) pada Program Studi Teknologi Informasi Fakultas Teknik
Universitas Muhammadiyah Palembang

Oleh:

Dheriessa Ayudhea
162022046

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH PALEMBANG
2026**

HALAMAN PENGESAHAN

PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENGELOMPOKKAN ANOMALI KOMENTAR NORMAL, SPAM, DAN BOT PADA PLATFORM TIKTOK



Oleh:

Dheriessa Ayudhea
162022046

Menyetujui,

Pembimbing Utama,

Jimmie, S. Kom., M. Kom
NBM/NIDN: 1340253/0222047702

Pembimbing Pendamping,

Dedi Haryanto, S. Kom., M. Kom
NBM/NIDN: 1337459/0201089001

Disetujui,
Dekan Fakultas Teknik

Ir. A. Junaidi, M.T
NBM/NIDN: 763050/0202026502

Program Studi Teknologi Informasi
Ketua Program Studi

Karnadi, S. Kom., M. Kom
NBM/NIDN: 1088893/0210038202

HALAMAN PENGESAHAN PENGUJI

Judul Skripsi : Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Anomali Komentar Normal, Spam, Dan Bot Pada Platform Tiktok

Oleh **Dheriessa Ayudhea** NIM 162022046 Skripsi ini telah disetujui dan disahkan oleh Tim Penguji Program Studi Teknologi Informasi Konsentrasi Rekayasa Perangkat Lunak Program Strata 1 Universitas Muhammadiyah Palembang pada 24 April.2026 dan telah Dinyatakan LULUS.

Palembang, 24 April 2026

Mengetahui,

**Ketua Program Studi Teknologi Informasi
Universitas Muhammadiyah Palembang**

**Tim Penguji
Ketua Penguji**



**Karnadi, S. Kom., M. Kom
NBM/NIDN: 1088893/0210038202**

**Jimmie, S. Kom., M. Kom
NBM/NIDN: 1340253/0222047702**

Penguji 1,

**Karnadi, S. Kom., M. Kom
NBM/NIDN: 1088893/0210038202**

Penguji 2,

**Kemas Muhammad Wahyu Hidayat, S.Kom., M.Kom
NBM/NIDN: 1255881/0225068904**

HALAMAN PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Dheriessa Ayudhea

Nim : 162022046

1. Karya tulis saya (Skripsi) merupakan sebuah karya asli serta belum pernah diajukan untuk memperoleh gelar akademik dengan baik (Sarjana) di Program Studi Teknologi Informasi Fakultas Teknik Universitas Muhammadiyah Palembang atau Perguruan Tinggi lainnya.
2. Karya tulis saya (Skripsi) tidak terdapat karya serta pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali serta tertulis dengan jelas dicantumkan nama pengarang serta memasukan kedalam daftar Pustaka.
3. Karya tulis saya (Skripsi) yang dihasilkan sudah melakukan pengecekan dengan keasliannya menggunakan plagirisme checker yang dipublikasikan melalui internet sehingga bisa diakses secara daring.
4. Dengan ini surat pernyataan yang saya buat secara sungguh-sungguh serta apabila terbukti terdapat penyimpangan serta ketidakbenaran dari pernyataan, maka saya bersedia menerima sanksi sesuai dengan peraturan serta perundang-undangan akademik Program Studi di Fakultas Universitas Muhammadiyah Palembang.

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 22 April 2026
Yang membuat pernyataan


METERAI
TEMPEL
22
22AAANX363006863
Dheriessa Ayudhea
Nim.162022046

MOTTO DAN PERSEMBAHAN

MOTTO

“Terus melangkah, meskipun perlahan, karena setiap usaha yang tidak dihentikan akan selalu menemukan jalannya”

PERSEMBAHAN

Puji dan syukur penulis panjatkan kepada Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Dengan penuh rasa terima kasih, karya sederhana ini penulis persembahkan kepada:

1. Untuk kedua orang tua-ku tercinta, yang selalu memberikan doa, kasih sayang, dukungan, serta pengorbanan yang tidak pernah berhenti. Terima kasih atas segala perjuangan, kesabaran, dan kepercayaan yang selalu diberikan kepada penulis. Tanpa doa dan dukungan dari kalian, penulis tidak akan mampu melewati setiap proses dan tantangan hingga akhirnya dapat menyelesaikan pendidikan ini.
2. Untuk adik-adik tercinta, yang selalu memberikan semangat, doa, dan keceriaan kepada penulis. Kehadiran kalian menjadi salah satu penyemangat bagi penulis untuk terus berusaha hingga dapat menyelesaikan pendidikan ini dengan sebaik mungkin.
3. Untuk diri sendiri, terima kasih karena telah berusaha, bertahan, dan terus melangkah meskipun prosesnya tidak selalu mudah. Terima kasih karena tetap kuat di tengah keraguan dan tidak menyerah hingga akhirnya bisa sampai pada tahap ini. Semoga ke depan tetap konsisten, terus berkembang, dan tidak berhenti berjuang meraih apa yang diimpikan.
4. Untuk dosen pembimbing, yang dengan sabar telah meluangkan waktu, tenaga, serta pikiran untuk membimbing penulis. Terima kasih atas arahan, saran, serta ilmu yang diberikan selama ini sehingga penulis dapat menyelesaikan skripsi ini dengan baik.

5. Untuk teman-teman seperjuangan yang selalu memberikan dukungan, bantuan, dan semangat selama masa perkuliahan hingga proses penyusunan skripsi. Terima kasih atas kebersamaan, motivasi, serta pengalaman yang telah dilalui bersama.

6. Untuk semua pihak yang telah membantu, baik secara langsung maupun tidak langsung, yang tidak dapat disebutkan satu per satu, namun telah memberikan kontribusi dan dukungan dalam proses penyelesaian skripsi ini.

Semoga karya sederhana ini dapat memberikan manfaat bagi pembaca serta menjadi langkah awal bagi penulis untuk terus belajar, berkembang, dan memberikan hal yang bermanfaat di masa yang akan datang.

ABSTRAK

Perkembangan pesat teknologi informasi dan komunikasi telah mendorong peningkatan penggunaan media sosial sebagai sarana utama bagi pengguna dalam menyampaikan opini publik. Dengan tingginya jumlah komentar pada platform TikTok, muncul beragam jenis komentar, termasuk komentar normal, *spam*, dan *bot* yang berpotensi menurunkan kredibilitas informasi. Maka dari itu, diperlukan penerapan metode yang efektif untuk melakukan klasterisasi terhadap komentar-komentar tersebut secara sistematis. Penelitian ini bertujuan untuk mengelompokkan komentar ke dalam kategori normal, *spam*, dan *bot* menggunakan pendekatan *unsupervised learning*. Metode yang digunakan adalah algoritma *K-Means Clustering* dengan tahapan prapemrosesan teks dan pembobotan kata menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)*. Data yang digunakan berupa komentar pengguna pada konten produk Skintific dan Glad2Glow di platform TikTok. Hasil penelitian menunjukkan bahwa penerapan algoritma *K-Means Clustering* terbukti efektif dalam membagi data komentar ke dalam tiga *cluster* dengan karakteristik yang signifikan berbeda, yaitu *cluster* komentar normal yang didominasi interaksi menengah dan kosakata beragam, *cluster spam* dengan interaksi tinggi dan variasi kata yang cukup, serta *cluster bot* yang ditandai dengan komentar singkat, repetitif, dan interaksi rendah. Secara keseluruhan, hasil analisis ini menyajikan wawasan mendalam terkait pola karakteristik komentar di media sosial. Penelitian ini diharapkan berkontribusi sebagai acuan praktis dalam pengembangan sistem moderasi yang lebih efektif serta pengembangan teknik analisis data teks berbasis klasterisasi.

Kata Kunci: *K-Means Clustering*; Komentar Spam; Akun Bot; Tiktok

ABSTRACT

The rapid development of information and communication technology has driven the increased use of social media as a primary means for users to express public opinion. With the high number of comments on the TikTok platform, various types of comments have emerged, including normal comments, spam, and bots that have the potential to reduce the credibility of information. Therefore, an effective method is needed to systematically cluster these comments. This study aims to group comments into normal, spam, and bot categories using an unsupervised learning approach. The method used is the K-Means Clustering algorithm with text preprocessing and word weighting stages using Term Frequency-Inverse Document Frequency (TF-IDF). The data used are user comments on Skintific and Glad2Glow product content on the TikTok platform. The results show that the application of the K-Means Clustering algorithm is effective in dividing comment data into three clusters with significantly different characteristics: a normal comment cluster dominated by medium interaction and diverse vocabulary, a spam cluster with high interaction and sufficient word variety, and a bot cluster characterized by short, repetitive comments and low interaction. Overall, the results of this analysis provide in-depth insights into the characteristic patterns of comments on social media. This research is expected to contribute as a practical reference in developing more effective moderation systems and developing clustering-based text data analysis techniques.

Keywords: *K-Means Clustering; Spam Comments; Bot Accounts; TikTok*

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Puji Syukur alhamdulillah saya panjatkan kepada kehadiran Allah SWT, atas berkat, nikmat dan karunia-Nya lah sehingga dapat menyelesaikan penyusunan laporan skripsi yang berjudul **“Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Anomali Komentar Normal, Spam, dan Bot Pada Platform Tiktok”**. Demikian penulisan laporan skripsi ini disusun untuk memenuhi salah satu persyaratan akademik pada Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Muhammadiyah Palembang

Selama penyusunan laporan ini, penulis menerima banyak dukungan dan arahan dari banyak orang yang membantu hingga laporan ini selesai. Oleh karena itu, penulis ingin mengungkapkan penghargaan kepada:

1. Bapak Prof. Dr. Abid Djazuli, S.E., M.M selaku rektor universitas Muhammadiyah Palembang.
2. Bapak Ir. A. Junaidi, M.T Selaku Dekan FT Universitas Muhammadiyah Palembang.
3. Bapak Karnadi, S.Kom., M.Kom Selaku Kepala Program Studi Teknologi Informasi.
4. Bapak Jimmie, S.Kom., M.Kom Selaku Dosen Pembimbing Utama.
5. Bapak Dedi Haryanto, S.Kom., M.Kom Selaku Dosen Pembimbing Pendamping.
6. Orang tua, Terima kasih atas doa, dukungan dan motivasi yang tiada henti.

Akhir kata, penulis mengucapkan terima kasih atas perhatian, dukungan, dan waktu yang telah diberikan.

Palembang, 06 April 2026



Dheriessa Ayudhea
162022046

DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PENGESAHAN PENGUJI	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRAK	vii
ABSTRACT	viii
KATA PENGANTAR	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
BAB I. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	5
1.3 Rumusan Masalah	5
1.4 Batasan Masalah.....	6
1.5 Tujuan Penelitian.....	6
1.6 Manfaat Penelitian	7
1.7 Sistematika Penulisan.....	8
BAB II. TINJAUAN PUSTAKA	10
2.1 Media sosial dan Tiktok	10
2.2 Komentar Normal, Spam, dan Akun Bot	12
2.3 Machine Learning	13
2.4 K-Means Clustering	14
2.5 Google Colab	17
2.6 Python	18
2.7 Penelitian Sebelumnya	20
BAB III. METODE PENELITIAN	23
3.1 Jenis Penelitian.....	23
3.2 Waktu dan Tempat Penelitian.....	23

3.3 Alat dan Bahan Penelitian	25
3.4 Sumber Data	26
3.5 Metode Pengumpulan Data	27
3.6 Kerangka Berpikir	27
3.7 Algoritma Pengolahan Data Yang Digunakan.....	31
BAB IV. HASIL DAN PEMBAHASAN	35
4.1 Deskripsi Dataset	35
4.2 Preprocessing Teks	36
4.3 Ekstraksi dan Normalisasi Fitur	39
4.4 Penerapan Algoritma K-Means	42
4.5 Evaluasi Model Clustering.....	45
4.6 Analisis Dan Interpretasi Hasil Clustering.....	48
4.7 Implikasi Hasil Penelitian	58
BAB V. KESIMPULAN DAN SARAN.....	60
5.1 Kesimpulan	60
5.2 Saran.....	60
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

Gambar 3.1 Kerangka Berpikir	28
Gambar 3.2 Tahapan Algoritma <i>K-Means Clustering</i>	32
Gambar 4.1 Tampilan Hasil Perolehan Data Komentar Pada <i>Apify Console</i> .	35
Gambar 4.2 Visualisasi Wordcloud	39
Gambar 4.3 Distribusi Jumlah Data pada Setiap Cluster	43
Gambar 4.4 Visualisasi Pola Cluster Menggunakan PCA.....	44
Gambar 4.5 Nilai <i>Silhouette Score</i> Hasil Clustering	46
Gambar 4.6 <i>Silhouette Plot</i> pada Setiap Cluster	47

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	21
Tabel 3.1 Jadwal Penelitian	23
Tabel 4.1 Cleaning Data	36
Tabel 4.2 case folding.....	36
Tabel 4.3 Slang Word	37
Tabel 4.4 Tokenizing	37
Tabel 4.5 Filtering	38
Tabel 4.6 Stemming Data	38
Tabel 4.7 Top 10 Kata Dominan pada Setiap Cluster.....	49
Tabel 4.8 Rata-rata Interaksi Komentar pada Setiap Cluster.....	50
Tabel 4.9 Rata-rata Panjang dan Pola Kata pada Setiap Cluster	51
Tabel 4.10 Rata-rata Nilai Cosine Similarity pada Setiap Cluster	52
Tabel 4.11 Ringkasan Hasil Clustering.....	54

BAB I

PENDAHULUAN

1.1 Latar Belakang

Transformasi perilaku manusia dalam berbagai aspek, mulai dari interaksi sosial, ranah profesional, hingga komunikasi antarpribadi, dipengaruhi secara signifikan oleh kemajuan teknologi informasi dan komunikasi. Kemajuan ini telah melahirkan era komputerasi, yang menguatkan penyebaran informasi secara kilat dan luas, melampaui batas-batas geografis[1]. Saat ini, teknologi informasi membawakan fungsi penting dalam memajukan pertumbuhan berbagai sektor, termasuk ekonomi, sosial, dan budaya, di berbagai negara. Dengan memahami tren dan arah perkembangan teknologi informasi, masyarakat dapat mempersiapkan diri serta mengambil langkah strategis untuk memanfaatkan potensi teknologi secara optimal. Peralihan pola interaksi masyarakat dari bentuk konvensional menuju komunikasi digital yang bersifat *real-time* melalui platform daring merupakan salah satu dampak spesifik dari pesatnya perkembangan TIK[2].

Media sosial merupakan salah satu manifestasi pemanfaatan teknologi informasi yang kini telah terintegrasi secara fungsional dalam kehidupan masyarakat digital. Media sosial telah bertransformasi menjadi platform multifungsi yang memfasilitasi interaksi sosial, pertukaran data, serta penguatan sektor promosi ekonomi digital, di samping fungsinya sebagai sarana komunikasi[3]. Melalui berbagai platform seperti Facebook, Instagram, YouTube, dan TikTok, pengguna dapat berinteraksi secara langsung, melampaui

batasan ruang dan waktu, serta berpartisipasi aktif dalam membangun jaringan sosial global. Penggunaan media sosial sudah memengaruhi pola komunikasi interpersonal orang dari yang bersifat tatap muka menjadi berbasis digital, menciptakan bentuk interaksi baru yang lebih cepat, terbuka, dan dinamis[4].

Di Indonesia, TikTok menjadi salah satu media sosial yang menunjukkan perkembangan paling pesat dibandingkan platform lainnya. Platform ini digunakan *user* untuk membuat dan berbagi video pendek kreatif, didukung oleh sistem rekomendasi berbasis algoritma yang secara mekanis menampilkan konten berdasarkan minat dan interaksi pengguna. Berdasarkan Laporan *Digital 2025: Indonesia*, yang diterbitkan oleh *DataReportal* bekerja sama dengan *We Are Social* dan *Meltwater* (2025), terlihat bahwa terdapat sekitar 108 juta pengguna TikTok aktif yang berusia di atas 18 tahun. Angka ini menunjukkan jika Indonesia ialah salah satu pasar TikTok terbesar di Asia Tenggara.[5]. Data tersebut menunjukkan bahwa TikTok sudah berkembang menjadi ruang sosial digital yang berpengaruh dalam cara masyarakat berkomunikasi, berinteraksi, dan mengonsumsi informasi. Namun, dengan tingginya volume interaksi pada platform tiktok dapat menimbulkan tantangan dalam hal moderasi konten ialah risiko munculnya komentar *spam* dan akun *bot*, yang dapat menurunkan kualitas interaksi dan kredibilitas informasi di platform tersebut.

Keberadaan komentar *spam* dan aktivitas akun *bot* di media sosial merupakan permasalahan krusial yang berpotensi mengganggu keamanan, keaslian informasi digital dan mengurangi kredibilitas konten. Komentar *spam*

sering dimanfaatkan untuk menyebarkan tautan promosi, penipuan, atau propaganda, sedangkan akun *bot* digunakan untuk mengotomatiskan aktivitas komentar, yang berpotensi mengganggu interaksi pengguna secara alami. Komentar *spam* biasanya menampilkan pola teks yang berulang, penggunaan emoji yang berlebihan, dan tautan eksternal yang mencurigakan[6]. Dalam penelitian ini, Pengambilan data komentar difokuskan pada konten produk dan iklan viral di platform TikTok, karena jenis konten tersebut memiliki tingkat interaksi tinggi dan sering menjadi target penyebaran komentar *spam* maupun aktivitas akun *bot*.

Dalam mengelompokkan komentar anomali atau *spam*, berbagai pendekatan telah dilakukan, baik menggunakan metode *supervised learning* maupun *unsupervised learning*. Pada pendekatan *supervised learning* terdiri dari *Naïve Bayes*, *Support Vector Machine (SVM)*, dan *Decision Tree*. Sementara itu, pada pendekatan *unsupervised learning*, algoritma seperti *K-Means Clustering*, *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)*, dan *Hierarchical Clustering*. Pendekatan *unsupervised learning* dianggap lebih fleksibel dalam mengolah data teks berjumlah besar, terutama ketika data pelatihan berlabel sulit didapatkan. Selain itu, algoritma seperti *K-Means* juga dikenal efisien dalam menemukan pola tersembunyi serta mengelompokkan komentar yang menunjukkan perilaku tidak wajar atau menyimpang dari pola umum pengguna. [7]. Algoritma ini telah digunakan dalam berbagai penelitian, seperti pendeteksian *spam* email, analisis sentimen, dan pengelompokan teks berbahasa Indonesia. Dengan penerapan pra-

pemrosesan teks yang tepat, algoritma ini berpotensi menjadi solusi efektif dalam pengelompokan komentar TikTok ke dalam kategori normal, *spam*, dan *bot*.

Dengan meningkatnya intensitas penggunaan platform Tiktok, dapat menyebabkan lonjakan besar pada jumlah komentar pengguna, baik yang bersifat normal, *spam* maupun yang dihasilkan secara otomatis oleh akun *bot* yang dapat menyebabkan penurunan kredibilitas platform. Oleh karena itu, diperlukan penerapan algoritma *K-Means Clustering* sebagai metode pengelompokan agar dapat menemukan dan menghubungkan pola komentar yang memiliki karakteristik normal, spam, dan bot pada platform TikTok berdasarkan kemiripan fitur, tanpa memerlukan data berlabel. Penelitian ini bertujuan untuk menganalisis karakteristik komentar anomali melalui pendekatan *unsupervised learning* sehingga diperoleh pemahaman pola komentar yang menyimpang dari interaksi normal pengguna. Diharapkan juga bahwa penelitian ini akan memberikan kontribusi ilmiah untuk kemajuan dalam analisis data teks, terutama pada komentar berbahasa Indonesia yang memiliki struktur linguistik dan variasi ekspresi yang beragam. Hasil penelitian ini diposisikan sebagai kajian konseptual dan eksperimental yang dapat dijadikan referensi dalam pengembangan sistem moderasi komentar serta sebagai dasar bagi penelitian lanjutan di bidang *text mining* dan analisis media sosial.

Oleh karena itu, berdasarkan permasalahan yang diuraikan diatas, penulis mengambil judul **“Penerapan Algoritma K-Means Clustering untuk Mengelompokkan Anomali Komentar Normal, Spam, dan Bot pada**

Platform TikTok”. Diharapkan dengan penggunaan metode ini mampu mengelompokkan komentar anomali seperti *spam* dan *bot* secara efisien serta memberikan pola dan karakteristik komentar anomali yang dihasilkan dari proses pengelompokan.

1.2 Identifikasi Masalah

Menelaah latar belakang permasalahan yang telah dikemukakan, identifikasi masalah dalam studi ini diformulasikan sebagai berikut:

1. Dengan tingginya volume interaksi pada platform tiktok dapat menimbulkan tantangan dalam hal moderasi konten ialah risiko munculnya komentar *spam* dan aktivitas akun *bot*, yang dapat menurunkan kualitas interaksi dan kredibilitas informasi di platform tersebut.
2. Komentar *spam* sering dimanfaatkan untuk menyebarkan tautan promosi, penipuan, atau propaganda, sedangkan akun *bot* digunakan untuk mengotomatiskan aktivitas komentar, yang berpotensi mengganggu interaksi pengguna secara alami.

1.3 Rumusan Masalah

Berdasarkan permasalahan tersebut, penelitian ini merumuskan pertanyaan utama yaitu sejauh mana efektivitas penggunaan algoritma *K-Means Clustering* dalam mengelompokkan komentar menjadi kelompok normal, *spam*, dan *bot* pada platform TikTok?

1.4 Batasan Masalah

Untuk memfokuskan kajian penelitian serta membatasi ruang lingkup pembahasan, ditetapkan batasan masalah sebagai berikut

1. Data penelitian dibatasi hanya 1.000 komentar publik saja yang diambil dari video konten Skintific vs Glad2Glow di platform TikTok.
2. Pengelompokan komentar difokuskan pada tiga kategori, yaitu komentar normal, *spam*, dan *bot*.
3. Penelitian ini tidak membahas tindakan lanjutan langsung ke dalam sistem resmi Tiktok, melainkan sebatas pada tahap pengelompokan dan analisis pola.
4. Analisis dilakukan pada komentar berbahasa Indonesia.

1.5 Tujuan Penelitian

Berdasarkan urgensi permasalahan tersebut, penelitian ini dilaksanakan dengan tujuan sebagai berikut:

1. Mengimplementasikan algoritma *K-Means Clustering* untuk melakukan pengelompokan komentar pada platform TikTok ke dalam kategori normal, *spam*, dan *bot* melalui kemiripan fitur.
2. Mengidentifikasi dan menganalisis pola serta karakteristik komentar *anomali* (*spam* dan *bot*) yang dihasilkan dari proses pengelompokan.

1.6 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi signifikan, baik secara teoretis maupun praktis, dengan rincian sebagai berikut:

1.6.1 Manfaat Teoretis

Manfaat teoretis dari penelitian ini mencakup kontribusi pada pengembangan teori *data mining*, terutama dalam mengasah metode analisis teks berbasis pengenalan pola. Hasil penelitian ini dapat berfungsi sebagai referensi empiris bagi peneliti selanjutnya yang mengkaji efektivitas algoritma *K-Means Clustering* dalam konteks media sosial. Lebih jauh, penelitian ini diharapkan memberikan kerangka pemikiran baru mengenai pola perilaku komentar, sehingga dapat mendukung klasifikasi anomali yang lebih akurat dalam membedakan komentar normal, *spam*, dan akun *bot*.

1.6.2 Manfaat Praktis

Secara praktis, hasil penelitian ini diharapkan dapat memberikan implikasi nyata bagi berbagai pihak, dengan rincian sebagai berikut:

1. Bagi Peneliti

Penelitian ini menjadi sarana bagi peneliti untuk memperdalam kompetensi teknis dalam pengimplementasian algoritma *K-Means Clustering*, serta memperluas pemahaman analitis mengenai karakteristik pola komentar pada platform media sosial.

2. Bagi Pengguna atau Kreator TikTok

Penelitian ini diharapkan dapat membantu pengguna dalam memahami jenis-jenis komentar yang muncul, sehingga dapat lebih bijak

dalam menanggapi komentar serta meningkatkan kualitas interaksi di platform.

3. Bagi Pengelola Platform atau Pengembang Sistem

Diharapkan penelitian ini akan digunakan sebagai referensi dalam proses pengembangan sistem moderasi komentar otomatis yang bertujuan untuk mengidentifikasi dan mengelompokkan komentar *spam* dan *bot*.

1.7 Sistematika Penulisan

Agar pembahasan dalam penelitian ini lebih terarah dan mudah dipahami, sistematika penulisan skripsi ini disusun dengan perincian sebagai berikut:

BAB 1 : PENDAHULUAN

Bab ini menguraikan poin-poin utama penelitian yang mencakup latar belakang permasalahan, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan laporan.

BAB II : TINJAUAN PUSTAKA

Bab ini menyajikan landasan teoretis yang relevan dengan topik penelitian. Materi disusun berdasarkan sintesis dari berbagai literatur ilmiah, mencakup buku, jurnal, serta artikel yang berkaitan langsung dengan pokok permasalahan yang diteliti

BAB III : METODELOGI PENELITIAN

Bab ini menguraikan tahapan prosedural penelitian serta metodologi yang diterapkan dalam proses pengelompokan komentar termasuk kategori normal, *spam*, dan *bot* pada platform TikTok.

BAB IV : HASIL DAN PEMBAHASAN

Pada bab ini memuat hasil dari pengelompokan anomali komentar normal, *spam*, dan *bot* yang dilakukan oleh peneliti terhadap komentar pada aplikasi Tiktok.

BAB V : PENUTUP

Bab terakhir ini memaparkan simpulan akhir atas hasil analisis yang telah dilakukan, serta memberikan saran-saran konstruktif bagi pengembangan aplikasi lebih lanjut.

DAFTAR PUSTAKA

- [1] M. W. Perdana, D. Haryanto, A. I. Alfresi, S. Hamidani, and A. F. Tegriansyah, "Sistem Informasi Himpunan Mahasiswa Berbasis Web Pada Prodi Teknologi Informasi Universitas Muhammadiyah Palembang," *J. Digit. Teknol. Inf.*, vol. 5, no. 1, p. 32, 2022, doi: 10.32502/digital.v5i1.4386.
- [2] K. G. Segara and M. I. P. Nasution, "Perkembangan Teknologi Informasi di Indonesia: Tantangan dan Peluang Peluang," *J. Sains Student Res.*, vol. 3, no. 1, pp. 21–33, 2025.
- [3] D. Macan, L. Bukit, A. Setiaji, T. Sutabri, K. Muhammad, and W. Hidayat, "RT / RW Menggunakan Metode Waterfall di," vol. 4, no. 4, 2023.
- [4] M. F. Swarna, A. Rumardani, E. adi Saputra, D. P. Nuryadi, M. D. Al-mufid, and N. Amalia, "Dampak Penggunaan Media Sosial Terhadap Pola Komunikasi Interpersonal," *Karimah Tauhid*, vol. 3, no. 1, pp. 1012–1019, 2024, doi: 10.30997/karimahtauhid.v3i1.11841.
- [5] "Digital 2025: Indonesia — DataReportal – Global Digital Insights."
- [6] Zena Lusi, Ayu Eka Saputri, and Tri Basuki Kurniawan, "Identifikasi Komentar Spam Pada Sosial Media," *Neptunus J. Ilmu Komput. Dan Teknol. Inf.*, vol. 2, no. 2, pp. 71–76, 2024, doi: 10.61132/neptunus.v2i2.100.
- [7] E. P. Laksono and A. Wicaksono, "Penyaringan Spam email menggunakan K-Means," *J. Spektro*, vol. 5, no. 2, pp. 26–32, 2022.
- [8] J. J. L. Tongkotow Liedfray, Fiony J. Waani, "Peran Media Sosial Dalam Mempererat Interaksi Antar Keluarga," *J. Ilm. Society*, vol. 2, no. 1, p. 5, 2022.
- [9] S. Sosrowidigdo, "Pengaruh Sosial Media dalam Mendukung Pertumbuhan Digitalisasi Ekonomi: Sebuah Kajian Literatur Berbantuan VOSviewer," *J. ISO J. Ilmu Sos. Polit. dan Hum.*, vol. 4, no. 2, p. 8, 2024, doi: 10.53697/iso.v4i2.2040.
- [10] Abdul Qadir, M. Ramli, "Media Sosial (Definisi, Sejarah Dan Jenis-Jenisnya)," *Al-Furqan J. Agama, Sos. dan Budaya*, vol. 3, no. 6, pp. 2713–2724, 2024.
- [11] I. Fadhilah, A. Azzahra, E. Haniammaria, D. Ardiyanti, and R. Ayu, "Pengaruh Aplikasi Tiktok Sebagai Media Promosi dan Kualitas Produk Terhadap Keputusan Pembelian Produk Moisturizer The Originote

memproduksi berbagai produk skincare yang disesuaikan dengan jenis kulit tertentu . Salah Menurut Infobrand . id (2023), Deb,” *J. Inov. Manajemen, Kewirausahaan, Bisnis dan Digit.*, no. 3, 2024.

- [12] R. Liswara, “Identifikasi akun bot pada media sosial twitter menggunakan metode extreme gradient boosting dan algoritma cuckoo search sebagai seleksi fitur,” *UIN Syarif Hidayatullah Jakarta Fak. Sains dan Teknol.*, 2022.
- [13] D. J. W. Ikram and S. F. Pane, “Deteksi Spam Bot pada Komentar Youtube : Tinjauan Literatur Sistematis,” *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 15, no. 2, pp. 103–123, 2023.
- [14] P. Chyan *et al.*, *Pengantar Machine Learning Pt. Mifandi Mandiri Digital*, vol. 1. 2024.
- [15] A. C. Muhammad *et al.*, “Dasar-dasar Pembelajaran Mesin,” p. 131, 2023.
- [16] N. Hendrastuty, “Pengertian Analisis Data,” *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024.
- [17] B. M. Al-Fahmi, “TA: Implementasi Data Mining Menggunakan K-Means Clustering dalam Mendukung Pengelompokan Tempat Wisata Kabupaten Bojonegoro,” 2023.
- [18] Y. Zouzou, “Deteksi kampanye pengikut palsu yang terkoordinasi tanpa pengawasan di media sosial,” 2024.
- [19] S. R. Noviah, “Mengenal Google Colab dan Manfaatnya,” *Exabytes*, 2023.
- [20] A. Hidayatullah and A. Berliana, “Workshop Pengenalan Dasar Pemrograman Python Dengan Google Colaboratory,” *Semin. Nas. Corisindo*, pp. 65–70, 2023.
- [21] S. O. F. M. P. E. Y. Indrawati, “Jurnal Pepadu,” *J. Pepadu*, vol. 6, no. 4, pp. 56–66, 2023.
- [22] S. Rahman *et al.*, *Python : Dasar Dan Pemrograman Berorientasi Objek Tahta Media Group*. 2023.
- [23] M. R. S. Alfarizi, M. Z. Al-farish, M. Taufiqurrahman, G. Ardiansah, and M. Elgar, “Use of Python as a Programming Language for Machine Learning and Deep Learning,” *Sci. Work Students Belief Monoth. (KARIMAH TAUHID)*, vol. 2, no. 1, pp. 1–6, 2023.
- [24] P. K. Berindikasi, M. A. Syawal, and B. A. Pradhana, “Implementasi Metode K-Means Untuk Implementation Of The K-Means Method For Grouping

- Words Indicative Of Cyberbullying In,” vol. 9, no. 2, pp. 330–347, 2025, doi: 10.52362/jisicom.v9i2.2163.
- [25] A. Muhariya, “Pengelompokan Komentar Pada Media Sosial Instagram Menggunakan Metode K-Means Clustering Untuk Identifikasi Awal Cyberbullying,” pp. 1–71, 2022.
- [26] M. Refa *et al.*, “Analisis Sentimen Terhadap Komentar Negatif (Hate Speech) Di Twitter Dengan Algoritma K-Means Clustering Menggunakan Rapidminer,” *J. Inf. Technol. Informatics Eng.*, vol. 1, no. 1, pp. 57–61, 2025.
- [27] M. A. Z. Larasati, N. A. S. Winarsih, M. S. Rohman, and G. W. Saraswati, “Penerapan Metode K-Means Clustering Dalam Menganalisis Sentimen Masyarakat Terhadap K-Poppers Pada Twitter,” *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 201, 2022, doi: 10.35889/progresif.v18i2.877.
- [28] N. Purnomo and W. Gata, “Pengelompokan Analisis Sentimen Komentar Youtube Terhadap Pengambilalihan Jalan Rusak di Lampung Menggunakan Algoritma Clustering,” *Progresif J. Ilm. Komput.*, vol. 20, no. 2, p. 701, 2024, doi: 10.35889/progresif.v20i2.1936.
- [29] A. Z. Syahputri, F. Della Fallenia, and R. Syafitri, “Kerangka berfikir,” *Tarb. J. Ilmu Pendidik. dan Pengajaran*, vol. 2, no. 1, pp. 160–166, 2023.
- [30] I. D. Sintawati and K. Mariskhana, “Analysis of Malnutrition Status in Toddlers Using the K-MEANS Algorithm Case Study in DKI Jakarta Province,” vol. 8, no. 4, pp. 2318–2324, 2024.