e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

IMPLEMENTASI METODE K-MEANS UNTUK PENGELOMPOKKAN KATA BERINDIKASI *CYBERBULLYING* PADA KOMENTAR TIKTOK

Implementation Of The K-Means Method For Grouping Words Indicative Of Cyberbullying In Tiktok Comment

Muhamad Akmal Syawal[1], Cahya Muhammad Rayhan[2], Bintang Arfiandi Pradhana [3], Yusuf Jordi Richardo[4], Khairul Rizal[5], Susliansyah[6]

Program Studi Teknologi Informasi[1,2,3,4,5],
Program Studi Sistem Informasi[6]
Fakultas Teknik dan Informatika[1,2,3,4,5,6]
Universitas Bina Sarana Informatika[1,2,3,4,5,6]

*Correspondent Author: 17230480@bsi.ac.id

Author Email: 17230968@bsi.ac.id [1], 17230480@bsi.ac.id[2], 17230759@bsi.ac.id[3], 17230487@bsi.ac.id [4], Khairul.krl@bsi.ac.id[5], susliansya.slx@bsi.ac.id[6]

Received: October 18,2025**. Revised:** November 20,2025**. Accepted:** December 04, 2025**. Issue Period:** Vol.9 No.2 (2025), Pp. 330-347

Abstrak: Perkembangan pesat media sosial seperti TikTok membawa peningkatan kasus *cyberbullying* yang berdampak negatif pada korban, termasuk tekanan mental. Penelitian ini bertujuan mengimplementasikan metode *clustering* sebagai pendekatan pembelajaran tanpa pengawasan untuk mengelompokkan kata-kata indikasi *cyberbullying* dalam komentar TikTok., Dataset komentar TikTok diolah melalui pra-pemrosesan teks seperti tokenisasi dan normalisasi. Metode *clustering* digunakan untuk mengelompokkan komentar berdasarkan kemiripan pola kata tanpa memerlukan data berlabel, Hasil pengelompokan mengidentifikasi pola kata konsisten yang menjadi indikasi *cyberbullying*, mendukung deteksi otomatis tindakan bullying di media sosial, Penelitian ini menyimpulkan bahwa pendekatan *clustering* efektif dalam mengenali ciri *cyberbullying* dan direkomendasikan untuk pengembangan sistem moderasi konten di TikTok guna mengurangi dampak negatif *cyberbullying*.

Kata kunci: TikTok, Pembelajaran Mesin, Perundungan Siber, Kecerdasan Buatan.

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

Abstract: *The rapid growth of social media platforms such as TikTok has led to an increase in cyberbullying cases, which negatively affect victims, including causing mental distress. This study aims to implement the clustering method as an unsupervised learning approach to group words that indicate cyberbullying in TikTok comments. The TikTok comment dataset is processed through text preprocessing techniques such as tokenization and normalization. The clustering method is then applied to group comments based on word pattern similarities without requiring labeled data. The clustering results identify consistent word patterns that serve as indicators of cyberbullying, supporting the automatic detection of bullying behavior on social media. This study concludes that the clustering approach is effective in recognizing characteristics of cyberbullying and is recommended for developing content moderation systems on TikTok to reduce the negative impact of cyberbullying.*

Keywords: *TikTok, Machine Learning, Cyberbullying, Artificial Intelligence*

I. PENDAHULUAN

Perkembangan media sosial saat ini sangat pesat dan telah menjadi bagian penting dalam kehidupan sehari-hari. Pengguna media sosial dapat dengan mudah melakukan interaksi kapan saja dan di mana saja. Salah satu platform media sosial yang kini sangat populer adalah TikTok. Platform ini memungkinkan penggunanya untuk berbagi video pendek sebagai sarana mengekspresikan diri dan menampilkan kreativitas. Namun, seperti halnya platform media sosial lainnya, TikTok juga rentan terhadap tindakan negatif seperti bullying, yang dalam bentuk digital disebut sebagai *cyberbullying*.

Cyberbullying ialah aksi penindasan di dunia maya yang dilakukan oleh suatu pihak kepada pihak lainnya dengan cara berkata kasar, menghina, mempermalukan maupun mengejek korban. Korban *cyberbullying* cenderung mendapatkan tekanan dan merasa stress hingga dapat mempengaruhi kesehatan mental. Banyak korban *cyberbullying* yang memilih untuk mengakhiri hidupnya sebab tidak sanggup menghadapi tekanan tersebut. Terdapat hubungan yang relevan antara sikap dari pelaku dan korban *cyberbullying*, yaitu semakin reaktif sikap pelaku maka semakin reaktif juga sikap korban *cyberbullying*. Hal tersebut menunjukkan begitu kuatnya dampak *cyberbullying* dalam kehidupan [1].

Karena maraknya kasus *cyberbullying* dan bahayanya yang signifikan, diperlukan metode yang efektif untuk mendeteksi dan mengelompokkan komentar berindikasi bullying. Salah satu pendekatan yang dapat digunakan adalah metode *clustering* dalam *unsupervised* learning, yang bertujuan menemukan kelompok kata atau komentar yang mengandung indikasi *cyberbullying* tanpa perlu data label sebelumnya. Salah satu tantangan utama dalam menangani *cyberbullying* adalah tingginya volume komentar di media sosial yang sulit dianalisis secara manual. Oleh karena itu, diperlukan pendekatan berbasis kecerdasan buatan (*Artificial Intelligence*) yang mampu mendeteksi pola-pola kata atau frasa yang berindikasi *cyberbullying* secara otomatis. Salah satu metode yang dapat digunakan untuk tujuan tersebut zadalah metode *clustering* dalam unsupervised learning[2].

Metode *clustering* berfungsi untuk menawarkan keunggulan berdasarkan kesamaan karakteristik tanpa memerlukan data berlabel dalam mengelompokkan data teks secara otomatis berdasarkan kemiripan antar-kata atau antar-komentar. Dengan menerapkan metode ini pada komentar TikTok, sehingga dapat membantu dalam identifikasi kata-kata berindikasi *cyberbullying* secara efisien diharapkan dapat diperoleh pengelompokan kata dan kalimat yang signifikan sebagai indikasi tindakan *cyberbullying*. Pendekatan ini sangat berguna mengingat banyaknya data komentar yang terus berkembang setiap saat di platform TikTok, Platform ini sangat dinamis dan terus bertambah setiap waktu.

Berdasarkan uraian di atas, penelitian ini fokus pada implementasi metode *clustering* untuk mengelompokkan kata yang berindikasi *cyberbullying* pada komentar di TikTok menggunakan teknik unsupervised learning. dan menggunakan alat bantu bahasa pemro-graman python yang dijalankan di aplikasi google collaboratory. Alasan pemilihan metode k-means,karena k-means merupakan suatu metode analisis data

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

yang melakukan proses pemodelan tanpa *supervise*(*unsupervised*) dan merupakan salah satu metode pengelompokan data dengan sistem partisi. Inti dari metode K-means adalah penentuan titik pusat k secara acak dan mempartisi data berdasarkan jarak antar data dan titik tengah [3].

Beberapa Hasil dari penelitian ini diharapkan dapat membantu dalam proses deteksi dini *cyberbullying* secara lebih efisien dan akurat pada media sosial TikTok. Dengan menerapkan metode *clustering*, diharapkan dapat dihasilkan pengelompokan kata atau komentar yang mencerminkan indikasi tindakan *cyberbullying*. Hasil dari penelitian ini dapat menjadi dasar pengembangan sistem moderasi konten otomatis di *platform* media sosial, khususnya TikTok, guna meminimalkan penyebaran ujaran kebencian dan prilaku negatif di dunia maya.

II. METODE DAN MATERI

Penelitian ini menggunakan pendekatan kuantitatif dengan metode unsupervised learning, khususnya metode *clustering*, untuk mengelompokkan kata atau komentar yang berindikasi *cyberbullying* pada platform media sosial TikTok. Pendekatan ini dipilih karena tidak memerlukan data berlabel (labelled data), sehingga dapat digunakan untuk menganalisis kumpulan data teks dalam jumlah besar secara otomatis.

Penelitian ini dilakukan melalui beberapa tahapan utama, yaitu pengumpulan data, prapemrosesan teks, ekstraksi fitur, penerapan metode *clustering*, serta evaluasi hasil pengelompokan. Berikut penjelasan masing-masing tahap. Data yang digunakan dalam penelitian ini berupa komentar pengguna TikTok yang diambil dari konten-konten publik menggunakan teknik *web scraping* atau API TikTok (jika tersedia). Komentar yang diambil difokuskan pada konten yang memiliki potensi mengandung interaksi negatif atau sensitif (misalnya perdebatan, ejekan, atau penghinaan).

Dataset yang diperoleh kemudian disimpan dalam format teks (.csv atau .txt) untuk diolah lebih lanjut. Jumlah data komentar yang digunakan diupayakan mencukupi agar hasil *clustering* memiliki distribusi yang representatif terhadap variasi bahasa dan gaya komunikasi di TikTok.

2.1. *Datamining*

Menurut Han, Kamber, dan Pei dalam bukunya yang berjudul *Data Mining: Concepts and Techniques*, K-Means adalah "sebuah algoritma *clustering* yang bekerja dengan cara meminimalkan jarak antara setiap data dengan *centroid* dari *cluster* yang sesuai". Dalam algoritma ini, centroid awal diinisialisasi secara acak dan kemudian diperbarui secara iteratif hingga mencapai konvergensi. Berbagai ragam tentang pendefinisian data mining, meliputi sebagai berikut [3]:

a. Penguraian (yang tidak sederhana) dari sekumpulan data menjadi informasi yang memiliki potensi secara implisit (tidak nyata/jelas) yang sebelumnya tidak diketahui.
b. Penggalian dan analisis, dengan menggunakan piranti otomatis atau semi otomatis, dari sejumlah besar data yang bertujuan untuk menemukan pola yang memiliki arti.
c. Data mining juga merupakan bagian dari *knowledge discovery* dalam *database* (KDD).

2.2. *Clustering*

Metode *clustering* dipilih sebagai pendekatan utama karena kemampuannya mengelompokkan data berdasarkan tingkat kesamaan karakteristik antar objek, dalam hal ini kata atau frasa pada komentar TikTok. Pendekatan ini memungkinkan deteksi dan pengelompokan kata-kata berpotensi mengandung unsur *cyberbullying* secara otomatis tanpa memerlukan pelabelan manual data. Metode *clustering* juga harus dapat mengukur kemampuannya dalam menemukan pola tersembunyi pada data melalui pengukuran kesamaan antar objek.[4] Salah satu metode pengukuran jarak yang umum dipakai adalah Weighted Euclidean Distance, yang menghitung jarak dua titik berdasarkan bobot atribut yang diberikan pengguna. Rumus umumnya adalah:

$$(\chi, \gamma) = (\sum_{k}^{n} \mu k \, (\chi_k - \gamma_k | r) \, 1/r \qquad (1)$$

Keterangan:
N = Jumlah *record* data
K = Urutan *field* data
r = 2

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

μk = bobot field yang diberikan oleh pengguna

Jarak adalah pendekatan yang umum dipakai untuk menetukan kesamaan atau ketidaksamaan dua vektor fitur yang dinyatakan dengan *ranking*. Apabila nilai *ranking* yang dihasilkan semakin kecil nilainya maka semakin dekat/tinggi kesamaan antara kedua vektor tersebut. Teknik pengukuran jarak dengan metode *Euclidean* menjadi salah satu metode yang paling umun digunakan. Pengukuran jarak dengan meotde *Euclidean* dapat dituliskan dengan persamaan berikut:

$$\text{Distance} = \sqrt{\sum_{k=1}^{N} \mu k \, (\chi_k - \gamma_k)^2} \qquad (2)$$

2.3. Algoritma K-means

Algoritma K-Means merupakan salah satu algoritma *clustering* berbasis metode partisi yang menggunakan titik pusat atau centroid sebagai referensi pengelompokan data. Dalam penerapannya, algoritma ini memerlukan tiga parameter yang seluruhnya ditentukan pengguna, yaitu jumlah cluster *k*, inisialisasi centroid awal, dan metode pengukuran jarak sistem. Biasanya, algoritma K-Means dijalankan beberapa kali dengan inisialisasi yang berbeda untuk menghindari hasil yang hanya terfokus pada local minima[5].
Penelitian ini mengaplikasikan algoritma K-Means untuk pengelompokkan kata-kata berindikasi *cyberbullying* dalam komentar pada platform TikTok. Hasil *clustering* memisahkan data ke dalam beberapa kategori, yakni:

- Non-*Cyberbullying*: kata-kata atau komentar yang tidak mengandung unsur bully.
- *Cyberbullying* Netral: kata yang berpotensi ambigu atau netral terkait unsur bully.
- *Cyberbullying* Ringan: kata-kata yang mengandung ejekan ringan atau sindiran.
- *Cyberbullying* Berat: kata-kata yang mengandung unsur intimidasi, ancaman, atau pelecehan serius.

Tahapan algoritma K-Means yang digunakan adalah:

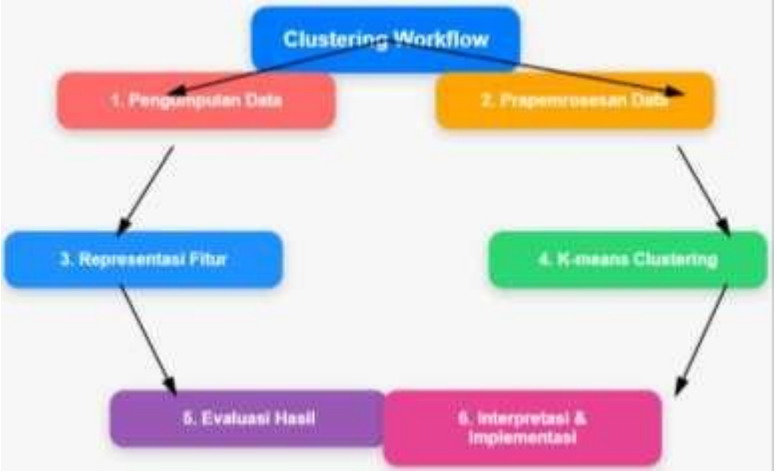
1. Menentukan jumlah cluster *k* sesuai dengan kategori pengelompokan yang diinginkan.
2. Menetapkan centroid awal secara acak dari data kata atau komentar.
3. Mengelompokkan setiap data ke centroid terdekat berdasarkan weighted Euclidean distance atau metrik jarak lain yang relevan.
4. Memperbarui posisi centroid berdasarkan rata-rata nilai data dalam cluster yang terbentuk.
5. Mengulangi langkah pengelompokan dan pembaruan centroid hingga centroid tidak berubah secara signifikan atau mencapai jumlah iterasi maksimal.

Pendekatan ini memungkinkan pengelompokan otomatis kata-kata yang berindikasi berbagai tingkat *cyberbullying* tanpa perlu data berlabel sebelumnya, sehingga efektif untuk menganalisis volume besar komentar TikTok secara kuantitatif[6].

III. PEMBAHASA DAN HASIL

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com



Gambar 1 Tahapan Penelitian

1.Pengumpulan Data : menggunakan web scraping (Apify) dan API TikTok untuk mendapatkan komentar dari konten publik, fokus pada komentar yang berpotensi mengandung *cyberbullying* (misalnya perdebatan, ejekan, penghinaan), data disimpan dalam format teks (.csv/.txt). 2.Prapemrosesan Data : tokenisasi untuk memecah teks menjadi kata-kata atau frasa, normalisasi untuk standarisasi teks agar konsisten, ekstraksi fitur untuk memilih kata/kata kunci yang relevan untuk analisis. 3.Representasi Fitur : Menggunakan teknik vektorisasi untuk merepresentasikan kata/komen sebagai vektor numerik, pemberian bobot (μk) sesuai relevansi fitur. 4.Penerapan Metode *Clustering* (K-means) : menentukan jumlah cluster (k), berdasarkan kategori *cyberbullying* yang diinginkan, menetapkan centroid awal secara acak, mengelompokkan data ke centroid terdekat menggunakan jarak Euclidean berbobot mengulang proses pengelompokan dan pembaruan centroid hingga konvergen atau maksimal iterasi tercapai. 5.Evaluasi Hasil *Clustering* : analisis pola kata dalam setiap cluster, meneliti apakah cluster merepresentasikan kategori *cyberbullying* : Non-*Cyberbullying*, *Cyberbullying* Netral *Cyberbullying* Ringan, *Cyberbullying* Berat. 6.Interpretasi dan Implementasi : mengidentifikasi kata-kata yang paling sering muncul dalam setiap cluster, mengintegrasikan hasil *clustering* dalam sistem moderasi otomatis di TikTok, menghasilkan laporan dan rekomendasi pengembangan sistem deteksi otomatis *cyberbullying*.

3.1. Pengelompokkan Data Komentar

Data dari 100 video dalam fyp yang diproses melalui web Apify sebagai data scraper untuk mencari comment yang berindikasi Positif/Netral/Negatif. Pada penelitian ini sambple yang diambil sebanyak 30 data pada console.apify.com

Tabel I. Data Komentar Positif/Netral/Negatif Dari Apify

Vid	Link Video	Komentar Positif	Komentar Netral	Komentar Negatif
1	https://vt.tiktok.com/ZSy1vwqNB/	3	2	95
2	https://vt.tiktok.com/ZSyYW2y1X/	34	11	55

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

3	https://vt.tiktok.com/ZSyYWagsw/	82	4	14
4	https://vt.tiktok.com/ZSyYwhJyc/	3	12	75
5	https://vt.tiktok.com/ZSyYwj32q/	5	2	93
6	https://vt.tiktok.com/ZSyYwSMQP/	6	7	87
7	https://vt.tiktok.com/ZSyYwu5sU/	9	34	57
8	https://vt.tiktok.com/ZSyYwcr5X/	2	43	55
9	https://vt.tiktok.com/ZSyYE2174/	3	5	92
10	https://vt.tiktok.com/ZSyYE2174/	7	12	81
11	https://vt.tiktok.com/ZSymS65vL/	2	10	88
12	https://vt.tiktok.com/ZSymSBwWj/	7	20	73
13	https://vt.tiktok.com/ZSymSudf2/	43	27	30
14	https://vt.tiktok.com/ZSymSuspW/	4	22	74
15	https://vt.tiktok.com/ZSymShBXX/	2	36	62
16	https://vt.tiktok.com/ZSymS34Hb/	0	8	92
17	https://vt.tiktok.com/ZSymSWtTg/	4	25	71
18	https://vt.tiktok.com/ZSymSQCo1/	1	12	87
19	https://vt.tiktok.com/ZSymSx9X2/	2	34	64
20	https://vt.tiktok.com/ZSymSEC56/	9	32	59
21	https://vt.tiktok.com/ZSymSgGwv/	6	15	79
22	https://vt.tiktok.com/ZSymSpjF6/	1	19	80
23	https://vt.tiktok.com/ZSymSG4Fy/	7	9	84
24	https://vt.tiktok.com/ZSymS3aft/	9	40	51
25	https://vt.tiktok.com/ZSymSKvJt/	2	5	93
26	https://vt.tiktok.com/ZSymSveMg/	0	9	91
27	https://vt.tiktok.com/ZSymS97QY/	30	21	49
28	https://vt.tiktok.com/ZSymSxSqt/	10	43	47
29	https://vt.tiktok.com/ZSymSQAp5/	1	1	98

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

30	https://vt.tiktok.com/ZSymSbVmE/	2	8	90

3.2. Penentuan Centroid awal

Dalam penentuan nilai *centroid* awal dilakukan secara acak pada data pembelian dengan terdiri dari tiga *clustering* yang akan digunakan untuk perhitungan iterasi ke 1 dan dapat dilihat pada tabel berikut:

Tabel II. Nilai *Centroid* Awal

Centroid	Komentar Positif	Komentar Netral	Komeentar Negatif
C1	6	7	87
C2	1	19	80
C3	2	8	90

3.3. Perhitungan Jarak Tiap Data ke Pusat Cluster Pada Iterasi Ke-1

Pengukuran jarak pada ruang jarak (distance space) Euclidean dengan menggunakan rumus dibawah ini [6]. Berikut ini merupakan salah satu contoh perhitungan jarak ke pusat cluster dengan menggunakan data T1 pada tabel.I dan data centroid awal C1pada tabel.II serta hasilnya ada di JC1 pada tabel.III sebagai berikut

$$d = \sqrt{(6-3)^2} + (7-2)^2 + (87 - 95)^2$$

$$= \sqrt{(3)^2} + (5)^2 + (-8)^2$$

$$= \sqrt{9^2} + 25^2 + 64^2$$

$$= \sqrt{98}$$

3.4 Menempatkan Data ke dalam Pusat *Cluster* Terdekat

Setelah dilakukan perhitungan berdasarkan salah satu contoh diatas, maka hasil perhitungan selengkapnya sebagai berikut :

Tabel III. Hasil Perhitungan Jarak Tiap Data ke Pusat Cluster Pada Iterasi Ke-1

Perhitungan Iterasi Ke-1						
Vid	JC1	JC2	JC3	C1	C2	C3
1	9,89949	22,75961	7,87401			*
2	42,70831	42,16634	47,52052		*	
3	105,42300	105,55567	110,41738	*		
4	13,34166	8,83176	15,55635		*	
5	7,87401	21,77154	7,34847			*
6	0	14,76489	5,09902	*		
7	40,47221	28,60070	42,59108		*	
8	48,33218	34,66989	49,49747		*	
9	6,16441	18,54724	3,74166			*
10	7,87401	9,27362	11,04536	*		
11	5,09902	12,08305	2,82843			*

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

12	19,13118	9,27362	21,39949		*	
13	70,83784	65,78754	75,11325		*	
14	19,94994	7,34847	21,35416		*	
15	38,50208	24,77907	39,59798		*	
16	7,87401	16,30951	2,82843			*
17	24,16613	11,22497	25,57342		*	
18	7,07107	9,89949	5,09902			*
19	35,69315	21,95449	36,77038		*	
20	37,65634	25,96151	39,82461		*	
21	11,31371	6,48074	13,63818		*	
22	14,76489	0	14,89966		*	
23	3,74166	12,32883	7,87401	*		
24	48,92851	36,68787	50,93133		*	
25	7,48331	19,13118	4,24264			*
26	8,71780	14,89966	2,44949			*
27	47,07440	42,49706	51,32252		*	
28	53,96295	41,78516	56,01785		*	
29	13,49074	25,45582	10,67708			*
30	5,09902	14,89966	0			*

3.5. Menentukan Centroid Baru dari Hasil Perhitungan Iterasi Ke-1

Mencari rata-rata dari hasil perhitungan iterasi ke-1 pada setiap Cluster Satu (C1), Cluster Dua (C2) dan Cluster Tiga (C3) untuk menentukan centroid baru dalam melakukan perhitungan iterasi ke-2. Adapun salah satu contoh perhitungan untuk mencari rata-rata pada Cluster Satu (C1) pada attribut
B.Masuk sebagai berikut:

$$= \frac{82 + 6 + 7 + 7}{4}$$

$$= \frac{102}{4}$$

$$= 25{,}2$$

Tabel IV. Hasil Rata-rata perhitungan untuk Cluster Satu (C1)

Cluster 1 (C1)			
Vid	Komentar Positif	Komentar Netral	Komentar Negatif
3	82	4	14
6	6	7	87
10	7	12	81
23	7	9	84
jml	jml	jml	jml
4	102	32	266
Rata-Rata	25,2	8	66,5

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

Tabel V. Hasil Rata-rata perhitungan untuk *Cluster* Dua (C2)

Cluster 2 (C2)			
Vid	Komentar Positif	Komentar Netral	Komentar Negatif
2	34	11	55
4	3	12	75
7	9	34	57
8	2	43	55
12	7	20	73
13	43	27	30
14	4	22	74
15	2	36	62
17	4	25	71
19	2	34	64
20	9	32	59
21	6	15	79
22	1	19	80
24	9	40	51
27	30	21	49
28	10	43	47
Jml	Jml	Jml	Jml
16	175	434	921
Rata- rata	10,9375	27,125	57,5625

Tabel VI. Hasil Rata-rata perhitungan untuk *Cluster* Tiga (C3)

Cluster 3 (C3)			
Vid	Komentar Positif	Komentar Netral	Komentar Negatif
1	3	2	95
5	5	2	93
9	3	5	92
11	2	10	88
16	0	8	92
18	1	12	87
25	2	5	93
26	0	9	91
29	1	1	98
30	2	8	90

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

Jml	Jml	Jml	Jml
10	19	62	919
Rata- rata	1,9	6,2	91,9

Dari hasil rata-rata pada tabel diatas dari setiap cluster maka di dapat centroid baru sebagai berikut:

Tabel VII. Centroid Baru

CENTROID	Komentar Positif	Komentar Netral	Komentar Negatif
C1	25,2	8	66,5
C2	10,9375	27,125	57,5625
C3	1,9	6,2	91,9

3.6. Perhitungan Jarak Tiap Data ke Pusat Cluster Pada Iterasi Ke-2

Hitung Euclidean distance dari semua data ke titik pusat yang baru pada tabel.VII, cara perhitungannya sama seperti yang telah dilakukan pada perhitungan iterasi ke-1. Setelah hasil perhitungan telah diperoleh, maka langkah selanjutnya adalah membandingkan hasil tersebut. Jika hasil posisi cluster pada iterasi ke-2 sama dengan posisi iterasi ke-1, maka proses dihentikan, namun jika tidak maka proses dilanjutkan ke iterasi ke-3.

Tabel VII. Hasil Perhitungan Jarak Tiap Data ke Pusat Cluster Pada Iterasi Ke-2

Perhitungan Iterasi 2						
Vid	JC1	JC2	JC3	C1	C2	C3
1	36,6208	45,9993	5,2783			*
2	14,7881	28,1751	48,9879	*		
3	77,4500	92,0137	111,8643	*		
4	24,1058	24,3746	18,0494			*
5	33,8570	44,9424	5,1981			*
6	28,1049	36,1035	6,4389			*
7	32,0732	7,1548	44,3786		*	
8	43,5372	18,1700	52,1864		*	
9	32,2099	42,3634	1,6309			*
10	23,7453	28,3870	13,3364			*
11	32,1168	35,7004	5,4461			*
12	22,7880	17,1610	23,4448		*	
13	42,2290	42,4840	81,6141	*		
14	25,8242	18,0350	24,1425		*	
15	36,7082	13,3653	42,2140		*	
16	35,8593	42,1997	2,6571			*
17	27,3695	15,1710	28,1684		*	
18	31,9889	34,5078	7,6694			*
19	35,9373	12,5865	39,3830		*	

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

20	29,8335	5,3003	42,3634		*	
21	23,4412	25,1853	16,0144			*
22	29,3920	26,2809	17,6033			*
23	25,4960	32,5334	9,9328			*
24	37,3991	14,5141	53,8169		*	
25	35,6355	43,4602	1,6309			*
26	35,4583	41,3725	3,3971			*
27	22,2731	21,1979	54,3016		*	
28	42,3632	19,6173	58,2379		*	
29	39,8584	49,1096	7,7756			*
30	33,0226	38,4608	2,6192			*

Tabel IX. Hasil Rata-rata perhitungan untuk *Cluster* Satu (C1)

Cluster 1 (C1)			
Vid	Komentar Positif	Komentar Netral	Komentar Negatif
2	34	11	55
3	82	4	14
13	43	27	30
Jml	jml	jml	jml
3	159	42	99
Rata-Rata	53	14	33

Tabel X. Hasil Rata-rata perhitungan untuk *Cluster* Dua (C2)

Cluster 2 (C2)			
Vid	Komentar Positif	Komentar Netral	Komentar Negatif
7	9	34	57
8	2	43	55
13	43	27	30
14	4	22	74
15	2	36	62
17	4	25	71
19	2	34	64
20	9	32	59
24	9	40	51
27	30	21	49
28	10	43	47
Jml	Jml	Jml	Jml
11	124	360	619

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

Rata- rata	11,27	32.727	56.27

Tabel XI. Hasil Rata-rata perhitungan untuk *Cluster* Tiga (C3)

Cluster 3 (C3)			
Vid	Komentar Positif	Komentar Netral	Komentar Negatif
1	3	2	95
4	3	12	75
5	5	2	93
6	6	7	87
9	3	5	92
10	7	12	81
11	2	10	88
16	0	8	92
18	1	12	87
21	6	15	79
22	1	19	80
23	7	9	84
25	2	5	93
26	0	9	91
29	1	1	98
30	2	8	90
Jml	Jml	Jml	Jml
16	49	136	1385
Rata- rata	3.0625	8.5	86.5625

Dari hasil rata-rata pada tabel diatas dari setiap cluster maka di dapat centroid baru sebagai berikut:

Tabel XII. Centroid Baru

CENTROID	Komentar Positif	Komentar Netral	Komentar Negatif
C1	53	14	33
C2	11,27	32.727	56.27
C3	3.0625	8.5	86.5625

3.7. Perhitungan Jarak Tiap Data ke Pusat Cluster Pada Iterasi Ke-3

Hitung Euclidean distance dari semua data ke titik pusat yang baru pada tabel.XII, cara perhitungannya sama seperti yang telah dilakukan pada perhitungan iterasi ke-2. Setelah hasil perhitungan telah diperoleh, maka langkah selanjutnya adalah membandingkan hasil tersebut. Jika hasil posisi cluster pada iterasi ke-3 sama dengan posisi iterasi ke-2, maka proses dihentikan, namun jika tidak maka proses dilanjutkan ke iterasi ke-4.

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

Tabel XIII. Hasil Perhitungan Jarak Tiap Data ke Pusat Cluster Pada Iterasi Ke-3

Perhitungan Iterasi 3						
Vid	JC1	JC2	JC3	C1	C2	C3
1	80.9197	49.7100	10.7854			*
2	29.5635	31.5185	44.3801	*		
3	35.5246	92.0147	107.4488	*		
4	65.3300	28.7252	12.1025			*
5	76.8375	47.3955	9.2573			*
6	72.0708	40.4545	3.3012			*
7	52.4609	2.7473	39.8550		*	
8	60.2163	13.8231	47.1153		*	
9	74.5788	45.7745	6.4710			*
10	66.5132	32.9795	7.4085			*
11	75.1132	40.6446	2.1199			*
12	61.2536	21.8981	18.0367			*
13	16.6733	41.6997	75.1970	*		
14	64.2028	22.2413	18.5445			*
15	60.2163	11.2680	36.8256		*	
16	79.2843	46.4442	6.3035			*
17	62.6578	17.1633	22.6914		*	
18	75.0733	39.2293	4.0785			*
19	59.5147	11.9965	34.3777		*	
20	52.5000	3.7090	37.2169		*	
21	65.7723	29.3877	10.4306			*
22	70.8378	29.9371	12.5728			*
23	68.8622	36.8200	4.6990			*
24	52.5000	9.0933	47.5729		*	
25	79.1265	46.8709	6.6157			*
26	78.5990	45.8725	5.3920			*
27	29.9000	23.3980	47.5130		*	
28	52.7826	13.9491	52.8078		*	
29	83.9999	53.2055	13.7103			*
30	76.8009	44.1706	3.6293			*

Tabel XIV. Hasil Rata-rata perhitungan untuk *Cluster* Satu (C1)

Cluster 1 (C1)			
Vid	Komentar Positif	Komentar Netral	Komentar Negatif

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

2	34	11	55
3	82	4	14
12	7	20	73
jml	jml	jml	jml
3	123	35	142
Rata-Rata	**41**	**11.67**	**47.33**

Tabel XV**.** Hasil Rata-rata perhitungan untuk *Cluster* Dua (C2)

Cluster 2 (C2)			
Vid	Komentar Positif	Komentar Netral	Komentar Negatif
7	9	34	57
8	2	43	55
15	2	36	62
17	4	25	71
19	2	34	64
20	9	32	59
24	9	40	51
27	30	21	49
28	10	43	47
Jml	Jml	Jml	Jml
9	77	308	515
Rata-rata	**8.56**	**34.22**	**57.22**

Tabel XVI. Hasil Rata-rata perhitungan untuk *Cluster* Tiga (C3)

Cluster 3 (C3)			
Vid	Komentar Positif	Komentar Netral	Komentar Negatif
1	3	2	95
4	3	12	75
5	5	2	93
6	6	7	87
9	3	5	92
10	7	12	81
11	2	10	88
12	7	20	73
14	4	22	74
16	0	8	92

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

18	1	12	87
21	6	15	79
22	1	19	89
23	7	9	84
25	2	5	93
26	0	9	91
29	1	1	98
30	2	8	90
Jml	Jml	Jml	Jml
18	60	163	1561
Rata-rata	3.33	9.056	86.72

Dari hasil rata-rata pada tabel diatas dari setiap cluster maka di dapat centroid baru sebagai berikut:

Tabel XVII. Centroid Baru

CENTROID	Komentar Positif	Komentar Netral	Komentar Negatif
C1	41	11.67	47.33
C2	8.56	34.22	57.22
C3	3.33	9.056	86.72

3.8. Perhitungan Jarak Tiap Data ke Pusat Cluster Pada Iterasi Ke-4

Hitung Euclidean distance dari semua data ke titik pusat yang baru pada tabel.XVII, cara perhitungannya sama seperti yang telah dilakukan pada perhitungan iterasi ke-3. Setelah hasil perhitungan telah diperoleh, maka langkah selanjutnya adalah membandingkan hasil tersebut. Jika hasil posisi cluster pada iterasi ke-4 sama dengan posisi iterasi ke-3, maka proses dihentikan, namun jika tidak maka proses dilanjutkan ke iterasi ke-5.

Tabel XVIII. Hasil Rata-rata perhitungan untuk *Cluster* Tiga (C4)

Perhitungan Iterasi 4						
Vid	JC1	JC2	JC3	C1	C2	C3
1	61.7246	49.7100	10.7854			*
2	29.5635	31.5185	44.3801	*		
3	35.5246	92.0147	107.4488	*		
4	65.3300	28.7252	12.1025			*
5	76.8375	47.3955	9.2573			*
6	72.0708	40.4545	3.3012			*
7	52.4609	2.7473	39.8550		*	
8	60.2163	13.8231	47.1153		*	
9	74.5788	45.7745	6.4710			*

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

10	66.5132	32.9795	7.4085			*
11	75.1132	40.6446	2.1199			*
12	61.2536	21.8981	18.0367			*
13	16.6733	41.6997	75.1970	*		
14	64.2028	22.2413	18.5445			*
15	60.2163	11.2680	36.8256		*	
16	79.2843	46.4442	6.3035			*
17	62.6578	17.1633	22.6914		*	
18	75.0733	39.2293	4.0785			*
19	59.5147	11.9965	34.3777		*	
20	52.5000	3.7090	37.2169		*	
21	65.7723	29.3877	10.4306			*
22	70.8378	29.9371	12.5728			*
23	68.8622	36.8200	4.6990			*
24	52.5000	9.0933	47.5729		*	
25	79.1265	46.8709	6.6157			*
26	78.5990	45.8725	5.3920			*
27	29.9000	23.3980	47.5130		*	
28	52.7826	13.9491	52.8078		*	
29	83.9999	53.2055	13.7103			*
30	76.8009	44.1706	3.6293			*

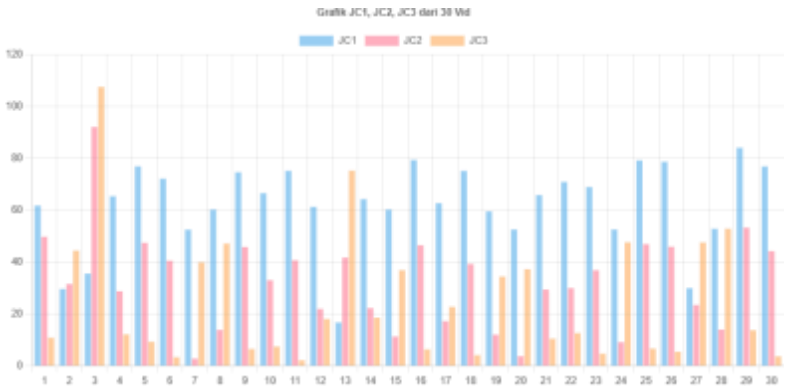
Karena pada iterasi ke-4 pada tabel XVII posisi *cluster* tidak berubah atau sama dengan posisi *cluster* pada iterasi ke-3 pada tabel.XII, maka proses perhitungan iterasi tidak dilanjutkan.

3.9. Ilustrasi

Pada gambar 2 menunjukkan bahwa komentar berindikasi *cyberbullying* menjadi kandidat utama mendapatkan grafik tertinggi. Berikut aplikasi wesite yang di gunakan untuk menentukan dan menjalankan menggunakan code html. 'W3Schools Tryit Editor' website ini mampu membuat grafik batang (bar chart) berdasarkan data pada Gambar 2: terdapat Tigapuluh video untuk menentukan komentar berindikasi *cyberbullying* dengan Nilai Trendah sampai tertinggi.

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com



Gambar 2. Grafik JC1, JC2, JC3 (Perhitungan Iterasi 4)

IV. KESIMPULAN

Kesimpulan penelitian ini menunjukkan bahwa metode *clustering* dalam pembelajaran tanpa pengawasan efektif dalam mengelompokkan kata-kata berindikasi *cyberbullying* pada komentar TikTok. Hasil pengelompokan tersebut mampu mengidentifikasi pola kata yang konsisten sebagai indikator *cyberbullying*, sehingga mendukung otomatisasi deteksi tindakan negatif di media sosial. Pendekatan ini sangat berguna dalam menghadapi volume besar komentar yang terus berkembang dan sulit dianalisis secara manual.

REFERENASI

[1] U. Khaira, R. Johanda, P. Eko, P. Utomo, and T. Suratno, "Sentiment Analysis Of *Cyberbullying* On Twitter Using SentiStrength," Indonesian Journal of Artificial Intelligence and Data Mining, vol. 3, no. 1, pp. 21–27, May 2020, doi: 10.24014/IJAIDM.V3I1.9145. Sentiment Analysis of Cyberbullying on Twitter Using SentiStrength - Neliti

[2] Sujatmiko, B. (2024). Analisis Metode *Clustering* dengan Algoritma DBSCAN Pengelompokan *Cyberbullying* di Instagram. Repository Universitas Medan Area.

[3] Maruti, E.S., et al. (2024). *Cyberbullying* Among Elementary School Students on TikTok: Analysis of Behavior Patterns. *FKIP UMADA Journal*, 2024.

[4] Oktaviyanti, A. (2025). Analisis Sentimen Ulasan Pelecehan di Media Sosial TikTok Menggunakan Algoritma K-Means *Clustering*. *Journal of Engineering*, University of Lampung.

DOI: 10.52362/jisicom.v9i2.2163

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.2 (December 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

[5] Oktaviyanti, A. (2025). Analisis Senitmen dan Pengelompokan Komentar *Cyberbullying* pada TikTok Menggunakan Algoritma K-Means *Clustering*. *Journal of Engineering*, University of Lampung.

[6] Erdiani, B.R.M. (2025). Novel Application of K-Means Algorithm for Sentiment Analysis on TikTok Comments. *Universitas Bumigora Journal*.

[7] Kartika, A.C.D. (2025). Bentuk dan Faktor *Cyberbullying* pada Kolom Komentar Konten Kreator TikTok. *Etheses UIN Malang*.

[8] Maruti, E.S. (2024). *Cyberbullying* Among Elementary School Students on Social Media: Classification and Symptoms Analysis. *FKIP UMADA Journal*.

[9] IEEE. (2025). Research on the Application of Weighted Distance K-Means *Clustering* Algorithm. *IEEE Xplore*. Scirp.org. (2025). Machine Learning for Identifying Harmful Online Behavior Using K-Means *Clustering*.

[11] Jurnal UNIMED. (2025). Algoritma K-Means *Clustering* untuk Mendiagnosis Bullying Berbasis Data Sosial Media.

[12] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," SMATIKA JURNAL, vol. 10, no. 02, pp. 71–76, Dec. 2020, doi: 10.32664/SMATIKA.V10I02.455.

[13] Rahman et al., "Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen," JEPIN (Jurnal Edukasi dan Penelitian Informatika), vol. 8, no. 2, pp. 299–303, Aug. 2022, doi: 10.26418/JP.V8I2.54704.

[14] M. A. Amrustian, W. Widayat, and A. M. Wirawan, "Analisis Sentimen Evaluasi Terhadap Pengajaran Dosen di Perguruan Tinggi Menggunakan Metode LSTM," JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 6, no. 1, pp. 535–541, Jan. 2022, doi: 10.30865/MIB.V6I1.3527.

[15] Y. yuli Astari, A. Afiyati, and S. W. Rozaqi, "Analisis Sentimen Multi-Class Pada Sosial Media Menggunakan Metode Long Short-Term Memory (LSTM)," Jurnal Linguistik Komputasional, vol. 4, no. 1, pp. 8–12, Apr. 2021, doi: 10.26418/JLK.V4I1.43.

DOI: 10.52362/jisicom.v9i2.2163