

Pengelompokan Pola Interaksi Pengguna Media Sosial Menggunakan Algoritma K-Means untuk Pemetaan Aktivitas Online

Rahma Yuni Simanullang*, Muhammad Iqbal

Fakultas Pasca Sarjana, Program Studi Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia

Email: ^{1,*}rahmayunisimanullang2009@gmail.com, ²muhammadiqbal@gmail.com

Email Correspondencial: rahmayunisimanullang2009@gmail.com

Abstrak—Penelitian ini bertujuan untuk mengelompokkan pola interaksi pengguna media sosial guna memperoleh pemetaan aktivitas online yang lebih terstruktur dan informatif. Aktivitas pengguna pada platform media sosial semakin meningkat dan beragam, mulai dari frekuensi unggahan, intensitas komentar, pola penyebaran informasi, hingga respons terhadap konten tertentu. Kompleksitas ini menimbulkan permasalahan dalam memahami karakteristik perilaku pengguna secara menyeluruh, terutama ketika data yang dihasilkan sangat besar dan tidak terstruktur. Untuk menjawab tantangan tersebut, penelitian ini menerapkan algoritma K-Means Clustering, salah satu metode data mining yang efektif dalam mengelompokkan data berdasarkan kesamaan karakteristik. Dataset yang digunakan berasal dari aktivitas pengguna yang mencakup jumlah postingan, jumlah likes, jumlah komentar, serta tingkat interaksi harian. K-Means digunakan untuk membagi data ke dalam beberapa klaster yang merepresentasikan tipe aktivitas pengguna, seperti pengguna aktif, semi-aktif, dan pasif. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu menghasilkan pemetaan aktivitas online yang jelas dan terukur, dengan nilai evaluasi menggunakan SSE dan silhouette score yang menunjukkan klaster terbentuk secara optimal. Dari proses pengelompokan tersebut diperoleh Cluster 1 sebagai pengguna pasif yang terdiri dari U001, U002, U005, U007, dan U009 dengan nilai centroid akhir C1 sebesar (0,13; 0,13; 0,10; 0,18), Cluster 2 sebagai pengguna aktif yang terdiri dari U003, U004, U006, dan U010 dengan nilai centroid akhir C2 sebesar (0,53; 0,55; 0,59; 0,49; 0,62), serta Cluster 3 sebagai pengguna sangat aktif yang hanya terdiri dari U008 dengan nilai centroid akhir C3 sebesar (1,00; 1,00; 1,00; 1,00).

Kata Kunci: Data Mining; Media Sosial; Algoritma K-Means; Pola Interaksi; Pemetaan Aktifitas Online

Abstract—This study aims to cluster social media user interaction patterns in order to obtain a more structured and informative mapping of online activities. User activity on social media platforms is increasing and diverse, ranging from the frequency of uploads, the intensity of comments, the pattern of information dissemination, to responses to certain content. This complexity creates problems in understanding the characteristics of user behavior as a whole, especially when the resulting data is very large and unstructured. To address this challenge, this study applies the K-Means Clustering algorithm, one of the data mining methods that is effective in clustering data based on similar characteristics. The dataset used comes from user activities that include the number of posts, the number of likes, the number of comments, and the level of daily interactions. K-Means is used to divide the data into several clusters that represent the types of user activities, such as active, semi-active, and passive users. The results show that the K-Means algorithm is able to produce a clear and measurable mapping of online activities, with evaluation values using SSE and silhouette scores indicating optimal cluster formation. From the grouping process, Cluster 1 was obtained as passive users consisting of U001, U002, U005, U007, and U009 with a final centroid value of C1 of (0.13; 0.13; 0.10; 0.18), Cluster 2 as active users consisting of U003, U004, U006, and U010 with a final centroid value of C2 of (0.53; 0.55; 0.59; 0.49; 0.62), and Cluster 3 as very active users consisting only of U008 with a final centroid value of C3 of (1.00; 1.00; 1.00; 1.00).

Keywords: Data Mining; Social Media; K-Means Algorithm; Interaction Patterns; Online Activity Mapping

1. PENDAHULUAN

Aktivitas pengguna di media sosial semakin berkembang pesat, tumbuh subur seiring dengan meningkatnya penetrasi dan penggunaan platform digital dalam kehidupan sehari-hari masyarakat global. Platform seperti Twitter (X), Instagram, Facebook, dan YouTube tidak lagi hanya berfungsi sebagai sarana komunikasi semata, melainkan telah bertransformasi menjadi representasi digital dari ruang publik, pasar, dan pusat informasi [1],[2]. Berbagai bentuk interaksi yang dilakukan pengguna seperti mengunggah konten baru, memberikan komentar yang mendalam atau singkat, memberikan tanda suka (likes) atau reaksi lainnya, serta membagikan ulang (sharing) informasi telah menjadi objek penting untuk dipelajari guna memahami perilaku pengguna secara menyeluruh. Lebih dari sekadar tindakan individual, pola interaksi kolektif ini merupakan cerminan nyata dari tren sosial yang sedang berlangsung, mengukur tingkat keterlibatan (engagement) komunitas, serta mengungkap dinamika komunikasi dan penyebaran informasi dalam komunitas online. Oleh karena itu, kemampuan untuk memahami dan mengurai pola interaksi pengguna media sosial bukan hanya sekadar akademis, tetapi telah menjadi suatu kebutuhan esensial dalam berbagai sektor yang mengandalkan analisis perilaku digital, mulai dari pemasaran, kebijakan publik, hingga pengembangan platform itu sendiri.

Namun, tantangan terbesar muncul dari besarnya volume data (Big Data) yang terus dihasilkan setiap detiknya di seluruh platform. Tingkat *velocity* (kecepatan), *volume* (jumlah), dan *variety* (variasi) data yang ekstrem menimbulkan permasalahan signifikan dalam proses identifikasi pola interaksi secara manual atau konvensional. Data yang sebagian besar bersifat tidak terstruktur, bervariasi jenisnya (teks, gambar, video, metadata interaksi), dan berjumlah masif membuat metode analisis tradisional, seperti *sampling* atau inspeksi visual, menjadi tidak lagi efektif dan tidak efisien. Diperlukan sebuah pendekatan yang mampu menyaring "kebisingan" data untuk menemukan "sinyal" pola yang berarti. Solusi yang relevan untuk mengatasi permasalahan kompleks ini adalah dengan memanfaatkan teknik data mining. Data mining, sebagai disiplin ilmu, memiliki kapabilitas unik untuk mengolah data dalam jumlah sangat besar (Big Data processing) dan secara sistematis menemukan pola tersembunyi, tren, dan asosiasi yang tidak terlihat secara langsung di dalamnya [3]. Dengan mengadopsi pendekatan berbasis komputasi ini, proses analisis dapat dilakukan lebih cepat, lebih

terstruktur, dan yang terpenting, memberikan hasil yang jauh lebih akurat dalam memahami karakteristik aktivitas dan segmentasi perilaku pengguna secara mendalam.

Data mining sendiri dapat didefinisikan sebagai proses ekstraksi pengetahuan (knowledge discovery) atau wawasan yang berguna dari data berukuran besar melalui penerapan serangkaian algoritma tertentu yang telah teruji. Dalam konteks penelitian ini, algoritma yang diusulkan untuk digunakan adalah K-Means Clustering [4],[5]. K-Means adalah metode klusterisasi non-hirarkis yang bekerja dengan mengelompokkan data berdasarkan kemiripan karakteristik intrinsik di antara *data point* tersebut. Prinsip dasarnya adalah membagi N objek data menjadi K kluster, di mana setiap objek data termasuk ke dalam kluster yang memiliki rata-rata (centroid) terdekat. K-Means dipilih secara spesifik dalam studi ini karena beberapa keunggulan strategis. Pertama, ia memiliki efisiensi komputasi yang tinggi, sehingga cocok untuk menangani *dataset* interaksi media sosial yang sangat besar. Kedua, kemampuannya dalam memetakan pola perilaku yang kompleks (berdasarkan fitur-fitur seperti frekuensi posting, rasio komentar/like, atau jam aktif) menjadi K kluster yang diskrit dan lebih mudah dipahami menjadikannya alat yang ideal. Dengan demikian, K-Means sangat cocok digunakan untuk tujuan utama studi ini: mengelompokkan tipe interaksi pengguna media sosial untuk segmentasi perilaku online [6].

Penggunaan klusterisasi dalam konteks perilaku pengguna media sosial bukanlah hal baru. Beberapa penelitian terkait telah dilakukan dalam beberapa tahun terakhir, menunjukkan tren peningkatan penggunaan *data mining* dalam analisis sosial. Sebagai contoh, Penelitian oleh Sari et al. (2022) membahas klusterisasi tingkat aktivitas pengguna Instagram. Menggunakan K-Means, mereka berhasil menemukan tiga pola utama perilaku pengguna, yang kemudian dikategorikan sebagai *Heavy User*, *Moderate User*, dan *Passive User*. Selanjutnya, studi oleh Nugroho dan Putri (2023) lebih jauh menerapkan K-Means untuk mengelompokkan frekuensi interaksi pengguna Twitter (X). Hasil penelitian mereka menghasilkan pembagian kluster yang diklaim akurat dalam mendukung analisis sentimen lanjutan, menunjukkan bahwa klusterisasi aktivitas dapat meningkatkan resolusi analisis. Selain itu, Penelitian oleh Wijaya (2023) memanfaatkan teknik *data mining* (meskipun mungkin mencakup lebih dari sekadar K-Means) untuk pemetaan topik diskusi pengguna Facebook dan secara eksplisit menunjukkan bahwa klusterisasi meningkatkan pemahaman terhadap distribusi perbincangan, mengidentifikasi kelompok pengguna berdasarkan minat topikal mereka. Yang paling mutakhir, penelitian oleh Rahman et al. (2024) bergerak ke arah *multimodal analysis*, menggabungkan K-Means dengan teknik pemrosesan bahasa alami (NLP) seperti TF-IDF dalam klusterisasi komentar YouTube. Kombinasi ini bertujuan untuk mengidentifikasi tren komunikasi digital berdasarkan kedalaman dan frekuensi interaksi teks.

Meskipun penelitian-penelitian sebelumnya telah meletakkan dasar yang kuat dalam klusterisasi aktivitas, Kontribusi fundamental penelitian ini terletak pada upaya untuk menghasilkan pemetaan aktivitas online yang lebih komprehensif dan multidimensi. Alih-alih hanya berfokus pada satu dimensi (misalnya, hanya frekuensi posting atau hanya sentimen), penelitian ini berupaya mengintegrasikan berbagai metrik interaksi (seperti rasio unggahan, komentar, like, *share*, dan waktu aktif) untuk membangun profil interaksi yang lebih lengkap. Melalui penerapan algoritma K-Means yang terstruktur pada fitur interaksi yang kaya ini, penelitian ini bertujuan memberikan gambaran segmentasi yang tajam tentang tipe pengguna berdasarkan tingkat dan jenis aktivitas mereka. Misalnya, mengidentifikasi kluster "Konten Kreator" (tinggi unggahan, tinggi interaksi masuk), "Komentator Aktif" (rendah unggahan, sangat tinggi komentar), dan "Penonton Pasif" (rendah semua metrik kecuali waktu tonton). Hasil pemetaan perilaku ini memiliki nilai aplikasi yang tinggi. Segmentasi ini dapat menjadi referensi vital bagi pengembang platform dalam merancang fitur baru yang sesuai dengan kebutuhan spesifik kluster pengguna; bagi peneliti sosial yang membutuhkan basis data terstruktur untuk studi perilaku; maupun bagi pihak lain (misalnya, *brand* atau pembuat kebijakan) yang membutuhkan analisis perilaku digital secara mendalam dan berbasis data yang kuat untuk strategi penargetan atau mitigasi informasi. Secara keseluruhan, penelitian ini bertujuan menjembatani kesenjangan antara volume Big Data media sosial dan kebutuhan akan pemahaman yang terstruktur mengenai dinamika interaksi pengguna.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dirancang secara sistematis melalui serangkaian tahapan yang terstruktur guna memastikan tercapainya tujuan utama, yaitu pemetaan pola interaksi pengguna media sosial. Kerangka kerja penelitian ini dimulai dari identifikasi masalah yang jelas hingga perumusan kesimpulan berdasarkan hasil klusterisasi data yang diperoleh. Berikut proses tahapan penelitian yg dilakukan penulis dan dapat dilihat pada Gambar 1 yg merupakan Diagram alurnya:

a. Identifikasi Masalah

Tahap awal ini merupakan penentuan fokus dan ruang lingkup penelitian. Ini melibatkan pengenalan kesenjangan pengetahuan atau tantangan yang perlu dipecahkan. Dalam konteks ini, masalah utamanya adalah tingginya volume dan kompleksitas data interaksi pengguna media sosial yang menyulitkan identifikasi pola perilaku secara manual, sehingga diperlukan metode komputasi untuk segmentasi perilaku (misalnya, siapa yang aktif, siapa yang pasif, dan bagaimana bentuk interaksi mereka).

b. Studi Literatur

Tahap ini melibatkan tinjauan kritis terhadap karya ilmiah, jurnal, dan penelitian terkait yang telah dipublikasikan sebelumnya. Tujuannya adalah untuk membangun kerangka teori yang kuat, memahami metode dan algoritma

(khususnya data mining dan K-Means) yang relevan, serta mengidentifikasi kontribusi unik dari penelitian ini dibandingkan dengan penelitian sebelumnya.

c. Pengumpulan Data

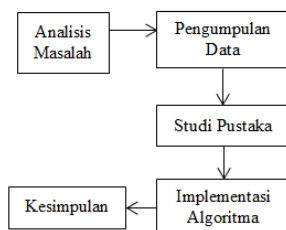
Fase ini berfokus pada akuisisi data mentah yang akan dianalisis. Data dikumpulkan dari platform media sosial tertentu (misalnya, Twitter, Instagram, atau Facebook) melalui API yang tersedia, mencakup metrik interaksi pengguna seperti frekuensi unggahan, jumlah likes, komentar, dan shares. Data mentah ini kemudian akan melalui proses pembersihan (data cleaning) dan prapemrosesan (preprocessing) agar siap digunakan dalam algoritma.

d. Penerapan Algoritma (K-Means Clustering)

Tahap inti penelitian adalah implementasi metode yang dipilih. Setelah data dipersiapkan, algoritma K-Means diterapkan. Proses ini mencakup penentuan jumlah kluster optimal (K), inisialisasi centroid, dan iterasi untuk mengelompokkan pengguna berdasarkan kemiripan pola interaksi mereka. Hasil dari tahap ini adalah terbentuknya kelompok-kelompok (kluster) pengguna yang merepresentasikan jenis aktivitas online yang berbeda.

e. Kesimpulan

Tahap akhir ini melibatkan interpretasi dan perumusan hasil yang diperoleh dari klusterisasi. Kesimpulan harus menjawab tujuan penelitian yang telah ditetapkan di awal dan merangkum temuan utama, seperti karakteristik dari setiap kluster pengguna yang berhasil diidentifikasi. Selain itu, tahap ini juga mencakup penyampaian saran atau implikasi praktis dari hasil penelitian untuk pengembangan platform atau analisis perilaku digital di masa mendatang.



Gambar 1. Tahapan Penelitian

2.2 Sosial Media

Media sosial adalah platform digital yang memungkinkan pengguna untuk membuat, membagikan, dan berinteraksi dengan konten secara daring [7]. Melalui media sosial, individu dapat berkomunikasi, bertukar informasi, membangun jejaring, serta mengekspresikan diri dalam berbagai bentuk seperti teks, gambar, video, dan siaran langsung. Media sosial juga memberikan ruang bagi komunitas untuk terbentuk dan berkembang berdasarkan minat, topik, atau hubungan sosial tertentu [8],[9]. Selain sebagai sarana komunikasi, media sosial kini berperan penting dalam berbagai bidang seperti pemasaran, pendidikan, bisnis, hiburan, hingga pengumpulan opini publik. Perkembangan media sosial yang cepat menjadikannya salah satu sumber data terbesar dalam memahami perilaku dan dinamika masyarakat modern [10],[11]

2.3 Algoritma K-Means

K-Means adalah sebuah algoritma clustering dalam data mining yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok (kluster) berdasarkan tingkat kemiripan antar data [12],[13],[14]. Algoritma ini bekerja dengan cara mencari pusat kluster (centroid) secara iteratif, lalu mengelompokkan setiap data ke centroid terdekat. Setelah semua data dikelompokkan, centroid akan diperbarui berdasarkan rata-rata posisi data dalam kluster tersebut, dan proses ini diulangi sampai posisi centroid stabil atau tidak berubah signifikan. K-Means efektif digunakan untuk menemukan pola tersembunyi dalam data berukuran besar dan membantu memetakan objek atau perilaku menjadi kelompok yang lebih terstruktur. Algoritma ini banyak digunakan dalam analisis perilaku pengguna, segmentasi pelanggan, analisis pola interaksi, hingga pengelompokan teks pada media sosial [15],[16],[17].

2.4 Data Mining

Data mining adalah proses penggalian dan analisis data dalam jumlah besar untuk menemukan pola, hubungan, tren, atau pengetahuan baru yang sebelumnya tidak diketahui secara eksplisit. Proses ini dilakukan dengan memanfaatkan teknik statistik, kecerdasan buatan, dan pembelajaran mesin guna mengekstraksi informasi yang bernilai dan bermanfaat sebagai dasar pengambilan keputusan. Data mining memungkinkan pengolahan data secara otomatis dan sistematis sehingga dapat mengungkap pola tersembunyi yang sulit ditemukan melalui analisis konvensional. Dalam konteks penelitian, data mining digunakan untuk membantu memahami karakteristik data serta memprediksi atau mengelompokkan objek berdasarkan kesamaan atribut tertentu [18],[19],[20].

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Tabel 1 menyajikan daftar atribut yang digunakan dalam penelitian ini sebagai variabel utama dalam proses analisis data. Setiap atribut merepresentasikan aktivitas pengguna pada platform media sosial yang berkaitan dengan tingkat interaksi

dan keterlibatan pengguna. Atribut Jumlah Posting menunjukkan total unggahan konten yang dibuat oleh pengguna, sedangkan Jumlah Like merepresentasikan total respons suka yang diterima atau diberikan pengguna. Selanjutnya, Jumlah Komentar menggambarkan tingkat interaksi pengguna dalam bentuk tanggapan teks, dan Jumlah Share menunjukkan seberapa sering konten pengguna dibagikan kepada pengguna lain. Atribut Durasi Aktif (menit) digunakan untuk mengukur total waktu aktivitas pengguna dalam menggunakan platform, yang mencerminkan intensitas keterlibatan pengguna secara keseluruhan. Atribut-atribut ini dipilih karena dinilai relevan dalam merepresentasikan pola perilaku pengguna yang menjadi dasar dalam proses analisis penelitian.

Tabel 1. Data Atribut

No	Atribut	Keterangan
1	Jumlah Posting	Total unggahan yang dibuat pengguna
2	Jumlah Like	Total like yang diberikan pengguna
3	Jumlah Komentar	Total komentar yang dibuat pengguna
4	Jumlah Share	Total konten yang dibagikan
5	Durasi Aktif (menit)	Total waktu aktif pengguna

Selanjutnya, Tabel 2 menampilkan data penelitian yang digunakan sebagai sampel dalam penelitian ini. Data tersebut berisi informasi aktivitas sepuluh pengguna yang diberi kode ID pengguna U001 hingga U010. Setiap baris data merepresentasikan satu pengguna, sedangkan setiap kolom menunjukkan nilai dari atribut yang telah didefinisikan pada Tabel 1. Data ini mencakup jumlah posting, jumlah like, jumlah komentar, jumlah share, serta durasi aktif pengguna dalam satuan menit. Dataset ini digunakan sebagai data awal untuk proses pengolahan dan analisis lebih lanjut, seperti normalisasi data, pemodelan, dan pengujian metode yang diterapkan dalam penelitian. Dengan adanya data ini, diharapkan dapat memberikan gambaran nyata mengenai variasi aktivitas pengguna yang dianalisis serta mendukung proses pengambilan kesimpulan secara objektif dan terukur.

Tabel 2. Data Penelitian

ID Pengguna	Jumlah Posting	Jumlah Like	Jumlah Komentar	Jumlah Share	Durasi Aktif (menit)
U001	5	120	30	10	450
U002	2	45	10	3	180
U003	15	300	80	40	900
U004	7	150	35	15	520
U005	1	25	5	1	120
U006	10	220	60	25	700
U007	3	60	15	5	250
U008	20	400	100	55	1100
U009	6	130	28	12	480
U010	12	260	70	30	800

3.2 Penerapan Algoritma K-Means

Setelah didapatkan data penelitian, maka selanjutnya akan dilakukan proses penerapan algoritma K-means. Yg dimana akan dilakukan proses dengan menggunakan 10 data.

Untuk Langkah Pertama : Melakukan Normalisasi Data (Min-Max)

$$\text{Rumus: } x_{\text{normalisasi}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Nilai minimum dan maksimum:

$$X_1 = \text{Posting : min} = 1, \text{ max} = 20$$

$$X_2 = \text{Like : min} = 25, \text{ max} = 400$$

$$X_3 = \text{Komentar : min} = 5, \text{ max} = 100$$

$$X_4 = \text{Share : min} = 1, \text{ max} = 55$$

$$X_5 = \text{Durasi : min} = 120, \text{ max} = 1100$$

U001 =

$$X_1 = \frac{5-1}{20-1} = \frac{4}{19} = 0,21$$

$$X_2 = \frac{120-25}{400-25} = \frac{95}{375} = 0,25$$

$$X_3 = \frac{30-5}{100-5} = \frac{25}{95} = 0,26$$

Perhitungan X4 hingga X5 sesuaikan dengan langkah sebelumnya . selanjutnya untuk again langkah perhitungan ID U002 Hingga U010 menyesuaikan dengan langkah ID U001.

Pada Tabel 3 menyajikan hasil proses normalisasi data yang telah dilakukan terhadap seluruh variabel penelitian, yaitu X1 hingga X5, untuk setiap objek dengan kode U001 sampai U010. Normalisasi data ini bertujuan untuk menyeragamkan skala nilai pada setiap variabel sehingga berada dalam rentang yang sama, yaitu antara 0 hingga 1. Dengan adanya proses normalisasi, perbedaan satuan dan rentang nilai antar variabel dapat diminimalkan, sehingga setiap variabel memiliki kontribusi yang seimbang dalam tahap analisis selanjutnya. Hasil normalisasi pada tabel ini menunjukkan variasi nilai untuk masing-masing objek penelitian, di mana nilai 0 merepresentasikan nilai terendah dan nilai 1 menunjukkan nilai tertinggi pada setiap variabel. Data yang telah dinormalisasi ini selanjutnya digunakan sebagai dasar dalam penerapan metode atau algoritma penelitian guna memperoleh hasil analisis yang lebih akurat dan objektif.

Tabel 3. Hasil Normalisasi data

ID	X1	X2	X3	X4	X5
U001	0,21	0,25	0,26	0,17	0,34
U002	0,05	0,05	0,05	0,04	0,06
U003	0,74	0,73	0,79	0,72	0,80
U004	0,32	0,33	0,32	0,26	0,41
U005	0,00	0,00	0,00	0,00	0,00
U006	0,47	0,52	0,58	0,44	0,59
U007	0,11	0,09	0,11	0,07	0,13
U008	1,00	1,00	1,00	1,00	1,00
U009	0,26	0,28	0,24	0,20	0,37
U010	0,58	0,63	0,68	0,54	0,69

Langkah kedua ialah menentukan Jumlah Cluster (K), yang dimana cluster yg dipilih ialah :

- Cluster 1 → Pengguna pasif
- Cluster 2 → Pengguna aktif
- Cluster 3 → Pengguna sangat aktif

Langkah Ketiga ialah menentukan Jumlah Centroid Awal :

Dipilih secara acak dari data Penelitian:

$$C1 = U002 \rightarrow (0,05; 0,05; 0,05; 0,04; 0,06)$$

$$C2 = U006 \rightarrow (0,47; 0,52; 0,58; 0,44; 0,59)$$

$$C3 = U008 \rightarrow (1,00; 1,00; 1,00; 1,00; 1,00)$$

Langkah Keempat ialah menghitung jarak Euclidean Distance dengan menggunakan rumus :

$$d = \sqrt{\sum_{i=1}^n (x_i - c_i)^2}$$

Perhitungan (U001):

Jarak U001 Ke C1

$$= \sqrt{(0,21 - 0,05)^2 + (0,25 - 0,05)^2 + (0,26 - 0,05)^2 + (0,17 - 0,04)^2 + (0,34 - 0,06)^2}$$

$$= \sqrt{0,026 + 0,040 + 0,044 + 0,017 + 0,078} = \sqrt{0,205} = 0,45$$

Jarak U001 Ke C2

$$= \sqrt{(0,21 - 0,47)^2 + (0,25 - 0,52)^2 + (0,26 - 0,58)^2 + (0,17 - 0,44)^2 + (0,34 - 0,0)^2}$$

$$= \sqrt{0,068 + 0,073 + 0,102 + 0,073 + 0,062} = \sqrt{0,378} = 0,61$$

Jarak U001 Ke C3

$$= \sqrt{(0,21 - 1,00)^2 + (0,25 - 1,00)^2 + (0,26 - 1,00)^2 + (0,17 - 1,00)^2 + (0,34 - 1,00)^2}$$

$$= \sqrt{3,066} = 1,75$$

U001 masuk Cluster 1

Selanjutnya untuk perhitungan U002 Hingga U010 Menyesuaikan langkah U001. Berdasarkan langkah penyelesaian keseluruhan jarak Euclidean Distance, maka didapatkan hasil Perhitungan jarak seluruh Data dan dapat dilihat pada Tabel 4 berikut:

Tabel 4. Hasil Perhitungan Jarak

ID	C1	C2	C3	Cluster
U001	0,45	0,61	1,75	C1
U002	0,00	0,95	2,12	C1
U003	1,42	0,47	0,54	C2
U004	0,58	0,42	1,52	C2
U005	0,10	1,07	2,23	C1
U006	0,95	0,00	1,01	C2
U007	0,18	0,85	1,98	C1
U008	2,12	1,01	0,00	C3
U009	0,49	0,56	1,65	C1
U010	1,11	0,22	0,63	C2

Langkah Kelima yaitu pembentukan Cluster sesuai dengan hasil perhitungan jarak Euclidean :

Cluster 1 (Pasif) → U001, U002, U005, U007, U009

Cluster 2 (Aktif) → U003, U004, U006, U010

Cluster 3 (Sangat Aktif) → U008

Berdasarkan Cluster yg sudah terbentuk karena proses pengelompokan pada iterasi pertama K-Means didasarkan pada jarak terdekat antara setiap data dengan centroid awal menggunakan nilai atribut yang telah dinormalisasi. Data U001, U002, U005, U007, dan U009 memiliki pola nilai yang relatif mirip dan cenderung rendah, sehingga jaraknya paling dekat dengan centroid Cluster 1 (Pasif). Data U003, U004, U006, dan U010 memiliki nilai atribut yang lebih tinggi dan seragam dibandingkan cluster pasif, sehingga lebih dekat dengan centroid Cluster 2 (Aktif). Sementara itu, data U008 memiliki nilai atribut yang paling tinggi dan berbeda secara signifikan dari data lainnya, sehingga jaraknya paling dekat dengan centroid Cluster 3 (Sangat Aktif) dan membentuk satu cluster tersendiri. Perbedaan tingkat kemiripan nilai inilah yang menyebabkan terbentuknya komposisi cluster tersebut pada iterasi pertama.

Langkah keenam yaitu Menghitung Centroid Baru :

Centroid Baru C1 :

$C1 = (0,13 ; 0,13 ; 0,13 ; 0,10 ; 0,18)$

Centroid Baru C2 :

$C2 = (0,53 ; 0,55 ; 0,59 ; 0,49 ; 0,62)$

Centroid Baru C3 :

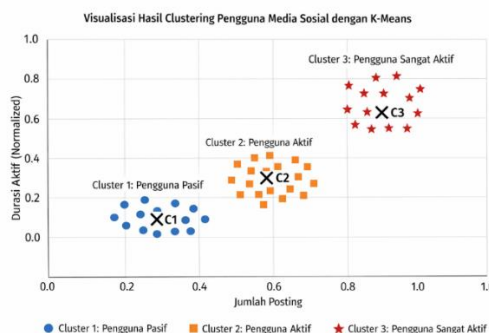
$C3 = (1,00 ; 1,00 ; 1,00 ; 1,00 ; 1,00)$

Iterasi dilanjutkan hingga tidak terjadi perubahan anggota cluster. Pada iterasi berikutnya, komposisi cluster tidak berubah, sehingga proses dinyatakan konvergen.

Berdasarkan hasil perhitungan manual algoritma K-Means menggunakan dataset aktivitas pengguna media sosial, diperoleh tiga cluster dengan karakteristik tingkat interaksi yang berbeda. Hasil ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan pengguna berdasarkan kesamaan pola aktivitas online secara efektif.

3.3 Visualisasi hasil clustering pengguna media sosial menggunakan algoritma K-Means

Pada Gambar 2 memperlihatkan pengelompokan pengguna media sosial berdasarkan dua variabel, yaitu Jumlah Posting pada sumbu X dan Durasi Aktif (Normalized) pada sumbu Y. Data pengguna dibagi menjadi tiga cluster yang ditandai dengan warna dan simbol berbeda. Cluster 1 (biru) merepresentasikan pengguna pasif, yaitu pengguna dengan jumlah posting rendah dan durasi aktif yang singkat. Cluster 2 (oranye) menunjukkan pengguna aktif, yaitu pengguna dengan jumlah posting dan durasi aktif pada tingkat menengah. Sementara itu, Cluster 3 (merah) menggambarkan pengguna sangat aktif, yaitu pengguna dengan jumlah posting tinggi serta durasi aktif yang lama. Titik XC1, XC2, dan XC3 merupakan centroid (pusat cluster) dari masing-masing kelompok, yang menunjukkan nilai rata-rata karakteristik pengguna dalam cluster tersebut. Secara keseluruhan, gambar ini menjelaskan bahwa algoritma K-Means berhasil mengelompokkan pengguna media sosial berdasarkan tingkat keaktifan mereka secara jelas dan terpisah.



Gambar 2. Visualisasi Hasil Clustering K-Means

4. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma K-Means pada data aktivitas pengguna media sosial mampu menghasilkan pengelompokan yang jelas dan informatif, sehingga mendukung tujuan penelitian dalam memetakan aktivitas online pengguna media sosial. Hasil penerapan algoritma K-Means Clustering pada data aktivitas pengguna media sosial dengan jumlah cluster sebanyak tiga ($K = 3$) dan menggunakan jarak Euclidean menunjukkan bahwa data berhasil dikelompokkan secara optimal berdasarkan tingkat aktivitas pengguna. Dari proses pengelompokan tersebut diperoleh Cluster 1 sebagai pengguna pasif yang terdiri dari U001, U002, U005, U007, dan U009 dengan nilai centroid akhir C1 sebesar (0,13; 0,13; 0,10; 0,18), Cluster 2 sebagai pengguna aktif yang terdiri dari U003, U004, U006, dan U010 dengan nilai centroid akhir C2 sebesar (0,53; 0,55; 0,59; 0,49; 0,62), serta Cluster 3 sebagai pengguna sangat aktif yang hanya terdiri dari U008 dengan nilai centroid akhir C3 sebesar (1,00; 1,00; 1,00; 1,00). Proses iterasi algoritma dihentikan setelah tidak terjadi perubahan anggota cluster pada iterasi selanjutnya, yang menandakan bahwa hasil clustering telah mencapai kondisi konvergen dan stabil, serta menunjukkan bahwa algoritma K-Means mampu memisahkan pengguna media sosial ke dalam kelompok dengan karakteristik aktivitas yang berbeda secara jelas dan terukur.

REFERENCES

- [1] S. Suryanto and W. Andriyani, "Sentiment Analysis of X Platform on Viral 'Fufufafa' Account Issue in Indonesia Using SVM," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 19, no. 1, p. 95, 2025, doi: 10.22146/ijccs.104158.
- [2] E. Edward, "Identifikasi Berita HOAX Berbasis Web Menggunakan Algoritma C4.5," *J. Ilmu Komput. dan Sist. Inf.*, vol. 9, no. 1, p. 53, 2021, doi: 10.24912/jiksi.v9i1.11558.
- [3] T. Firmansyah, P. Poningsih, and S. R. Andani, "Analisis Clustering Algoritma K-Means Sebagai Rekomendasi Penambahan Koleksi Buku Di Perpustakaan Madrasah Tsanawiyah Negeri 2 Simalungun," *Zahra Bull. Big data, Data Sci. Artif. Intell.*, vol. 1, no. 1, pp. 44–48, 2022.
- [4] L. Al-Alawi, J. Al Shaqsi, A. Tarhini, and A. S. Al-Busaidi, "Using machine learning to predict factors affecting academic performance: the case of college students on academic probation," *Educ. Inf. Technol.*, vol. 28, no. 10, pp. 12407–12432, 2023, doi: 10.1007/s10639-023-11700-0.
- [5] L. Erlistya, P. S. Sumarna, and ..., "Implementasi Algoritma K-Means Clustering pada Studi Kasus Seleksi Mahasiswa Berprestasi menggunakan Visual Basic," *Pros. Semin.*, vol. 4, no. 2721, 2021.
- [6] H. Ani, D. Nofriansyah, and I. Mariami, "Implementasi Data Mining Untuk Pengelempokan Buku Di Perpustakaan Yayasan Nurul Islam Indonesia Baru Dengan Metode K-Means Clustering," *J. Cyber Tech*, vol. 1, no. 1, 2021.
- [7] S. Hariyanti, "Pemanfaatan Media Sosial Tiktok Sebagai Sarana Promosi Kripik Jamur Tiram," *Al-Muraqabah J. Manag. Sharia Bus.*, vol. 2, no. 2, pp. 126–145, 2022.
- [8] F. Saefulloh, "Analisis Sentimen Masyarakat Terhadap E-Commerce Pada Media Sosial Twitter Menggunakan Metode Support Vector Machine (Svm) (Studi Kasus Shopee Dan Tokopedia)," *SI thesis, Univ. Muhammadiyah Purwokerto*, vol. 8, no. 2502, pp. 79–88, 2023.
- [9] R. S. Wulandari and F. K. Sari, "Media Sosial sebagai Platform Pembelajaran Alternatif di Era Digital," *Pros. Nas. Pendidik. LPPM IKIP PGRI Bojonegoro*, vol. 3, no. 1, pp. 279–288, 2023.
- [10] A. S. Sembiring and T. S. Alasi, "Penerapan Data Mining Menggunakan Algoritma Apriori Pada Peminjaman Buku di Perpustakaan Pada Pesantren Babul Ulum," *J. Armada Inform.*, vol. 7, no. 2, pp. 323–327, 2023.
- [11] T. P. Lestari, "Analisis Text Mining pada Sosial Media Twitter Menggunakan Metode Support Vector Machine (SVM) dan Social Network Analysis (SNA)," *J. Inform. Ekon. Bisnis*, vol. 4, no. 3, pp. 65–71, 2022, doi: 10.37034/infv4i3.146.
- [12] I. Kanedi and E. Suryana, "Penerapan Metode K-Means Clustering Dalam Pengelompokan Data Pasien Rawat Inap Peserta BPJS Di Rumah Sakit Umum Daerah Kabupaten Kaur," vol. 20, no. 2, pp. 493–500, 2024.
- [13] K. D. A. N. Knn, "Analisis clustering pada pemilihan tema vs ga dengan metode k-means dan knn clustering analysis on vs ga theme selection using k-means and knn methods," *Komputika J. Sist. Komput.*, vol. 7, no. 3, 2024.
- [14] C. Penerima, B. Kip, and S. Usanto, "Penerapan Data Mining Dengan Mengimplementasikan Algoritma K-Means Dalam Proses Clustering Untuk Pengelompokan Mahasiswa," *Buuld. Informatics, Technol. Sci.*, vol. 5, no. 1, pp. 70–85, 2023, doi: 10.47065/bits.v5i1.3411.
- [15] D. Aulia, M. Safii, and D. Suhendro, "Penerapan Algoritma K-Means dalam Proses Clustering Penilaian Kinerja Aparatur Sipil Negera di Sekretariat DPRD Pematangsiantar," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 6, no. 1, p. 47, 2021, doi: 10.30645/jurasik.v6i1.270.
- [16] M. Y. Matdoan, "Perbandingan Algoritma K-Means Dan K-Medoids Dalam Pengelompokan Tingkat Kebahagiaan Provinsi Di Indonesia," *BULLET J. Multidisiplin Ilmu*, vol. 1, no. 05, pp. 830–839, 2022.
- [17] R. Riadi and Mesran, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Analisa Penjualan Parfume," *J. Informatics, Electr. Electron. Eng.*, vol. 2, no. 4, pp. 138–145, 2023, doi: 10.47065/jieec.v2i4.1181.
- [18] A. T. Basalamah and R. Setyadi, "Penerapan Algoritma K-Means Clustering Pada Tingkat Penyelesaian Pendidikan Di Provinsi Indonesia," *J. Inform. dan Teknol. Komput.*, vol. 4, no. 2, pp. 114–121, 2023.
- [19] A. Supriyadi, A. Triayudi, and I. D. Sholihati, "Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 6, no. 2, pp. 229–240, 2021, doi: 10.29100/jupi.v6i2.2008.
- [20] M. Afrizal, I. Saputra, R. Satria, and Rahmaddeni, "Analisis Performa Algoritma K-Means Clustering untuk Segmentasi Pasar di UMKM Muhammad," *VISA J. Visions Ideas*, vol. 5, no. 2, pp. 647–657, 2025.