



BAB III METODE PENELITIAN

Bab ini menguraikan prosedur penelitian dengan pendekatan metode campuran (*mixed methods*) berbasis kerangka kerja *Knowledge Discovery in Databases* (KDD). Aspek kuantitatif diterapkan melalui komparasi algoritma *K-Means* dan *K-Medoids* (PAM) yang divalidasi menggunakan *Silhouette Coefficient* untuk menentukan pemodelan segmentasi kinerja konten dengan akurasi paling optimal. Sementara itu, aspek kualitatif menggunakan analisis konten berbasis model AIDA untuk menginterpretasikan karakteristik kluster terbaik serta merumuskan rekomendasi strategi pemasaran. Sebagai luaran aplikatif, hasil pemodelan ditransformasikan ke dalam prototipe lembar kerja digital interaktif sebagai instrumen evaluasi performa konten berkala bagi UMKM X.

3.1 Desain Penelitian dan Definisi Operasional Variabel

Penelitian ini menerapkan pendekatan metode gabungan (*mixed methods*) dengan model sekuensial. Pada tahap awal, pendekatan kuantitatif digunakan sebagai metode utama untuk mengolah data numerik dan mentransformasikan karakteristik konten kualitatif menjadi data biner agar dapat dikelompokkan secara objektif melalui studi komparasi algoritma *K-Means* dan *K-Medoids* (PAM) yang divalidasi dengan *Silhouette Coefficient*. Selanjutnya, pendekatan kualitatif diterapkan melalui analisis isi berbasis model hierarki komunikasi AIDA (*Attention, Interest, Desire, Action*) untuk menginterpretasikan makna serta karakteristik pesan di setiap kluster performa dari algoritma terbaik yang terbentuk



Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.

2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.

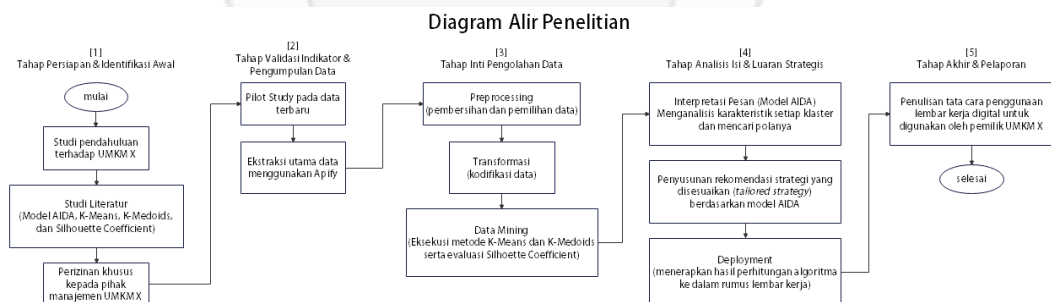
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Universitas Islam Indragiri

Secara umum, alur pelaksanaan penelitian ini dirancang secara sistematis ke dalam lima tahapan utama, yaitu:

1. Tahap persiapan dan identifikasi awal,
2. Tahap validasi indikator dan pengumpulan data melalui *pilot study* serta *web scraping*,
3. Tahap inti pengolahan data menggunakan siklus KDD (*Preprocessing, Transformation*, serta *Data Mining* komparatif *K-Means* vs *K-Medoids* dan evaluasi *Silhouette Coefficient*),
4. Tahap analisis isi menggunakan model AIDA serta luaran strategis berupa dokumen rekomendasi dan *deployment* formula, hingga
5. Tahap akhir dan pelaporan.

Untuk memberikan gambaran komprehensif mengenai jalannya penelitian, seluruh prosedur pelaksanaan studi tersebut dirangkum dalam diagram alir Gambar 3.1 berikut.



Gambar 3.1 Diagram Alir Keseluruhan Penelitian

Berdasarkan Gambar 3.1, diagram alir tersebut menegaskan keterkaitan antara proses komputasi data dan interpretasi manajerial. Melalui lima fase yang terstruktur, alur ini menjamin transparansi metodologi, mulai dari pemrosesan 188 data unggahan UMKM X di Orange Data Mining untuk membandingkan performa *K-Means* dan *K-Medoids* berbasis *Silhouette Coefficient*, hingga hasil pemodelan

dari algoritma paling optimal ditransformasikan ke dalam prototipe lembar kerja digital (Excel). Seluruh rangkaian prosedur ini membentuk satu kesatuan sistematis untuk menghasilkan rekomendasi strategi pemasaran konten yang valid dan aplikatif.

3.1.1 Subjek, Objek, dan Justifikasi Pemilihan Data

Subjek penelitian ini adalah akun Instagram resmi milik UMKM X, sebuah pelaku usaha kuliner di Kota Tembilahan yang dipilih melalui teknik *purposive sampling*. Pemilihan UMKM X dibanding pelaku usaha lainnya didasarkan pada pertimbangan bahwa akun ini merupakan salah satu pelopor usaha kuliner lokal yang aktif memanfaatkan media sosial secara konsisten, namun masih menghadapi kendala efektivitas interaksi audiens. Pemilihan UMKM skala lokal ini sejalan dengan urgensi adopsi teknologi data demi meningkatkan daya saing sektor usaha kecil sebagaimana ditekankan oleh Kedi dkk. [8]. Guna menjaga etika penelitian dan kerahasiaan data bisnis, identitas asli akun disamarkan menjadi UMKM X.

Adapun objek penelitian berupa data log publik historis unggahan yang mencakup metrik interaksi (*likes* dan *comments*) serta karakteristik komponen pesan. Penentuan objek riset serta pendekatan eksploratif ini didasarkan pada beberapa pertimbangan utama berikut:

1. Hasil Studi Pilot, evaluasi awal terhadap 20 unggahan terbaru menunjukkan bahwa meskipun mayoritas konten telah menyertakan elemen visual dan instruksi langsung, terdapat celah interaksi (*engagement gap*) yang signifikan serta pola publikasi yang belum konsisten. Data dari hasil *pilot study* tersebut disajikan pada Gambar 3.2.



no	post_date	post_days_gap	business_tagged	sales/promo_language	CTA_present	transaction_link	likes	comments	classification	type	captions
1	2/4/2026	22	yes	yes	yes	yes	26	0	promotional	carousel	Hi guys y
2	11/3/2026	1	yes	yes	yes	no	49	0	promotional	video	Bukber j
3	10/3/2026	3	yes	yes	yes	yes	61	0	promotional	carousel	Paket Bul
4	7/3/2026		yes	yes	yes	yes	38	2	promotional	carousel	Yang mau
5	7/3/2026	1	yes	yes	yes	no	14	1	promotional	carousel	Salah si
6	6/3/2026	2	yes	yes	yes	yes	111	1	promotional	carousel	Salah si
7	4/3/2026		yes	yes	yes	yes	32	1	promotional	carousel	Yang mau
8	4/3/2026	6	yes	yes	yes	yes	22	1	promotional	carousel	Pesan h
9	26/2/2026	1	yes	yes	yes	yes	17	1	promotional	carousel	Rekome
10	25/2/2026		yes	yes	yes	yes	94	1	promotional	carousel	Paket R
11	25/2/2026	2	yes	yes	yes	yes	53	1	promotional	carousel	TEH TA
12	23/2/2026		yes	yes	yes	yes	17	1	promotional	carousel	Paket Bul
13	23/2/2026	1	yes	yes	yes	yes	84	1	promotional	carousel	PAKET
14	22/2/2026		yes	yes	yes	yes	39	1	promotional	carousel	Hamper
15	22/2/2026	1	yes	yes	yes	yes	7	1	promotional	carousel	Cake tir
16	21/2/2026		yes	yes	yes	yes	89	1	promotional	carousel	Cemilar
17	21/2/2026		yes	yes	yes	yes	45	1	promotional	carousel	Yang me
18	21/2/2026	1	yes	yes	yes	yes	11	1	promotional	carousel	Paket Bul
19	20/2/226	1	yes	yes	yes	yes	43	3	promotional	carousel	Paket bi
20	19/2/2026		yes	yes	no	no	68	1	possibly	carousel	

Gambar 3.2 Data 20 Unggahan Study Pilot

- Anomali interaksi data, berdasarkan Gambar 3.2, terdapat fluktuasi interaksi yang tidak seimbang pada konten tertentu. Anomali ini terlihat jelas pada konten nomor 5 (113 *likes*) dan nomor 10 (96 *likes*) yang melonjak jauh di atas rata-rata keseluruhan (49 *likes* dan 1 *comment*). Keberadaan nilai pencilan (*outliers*) yang ekstrem ini menjadi justifikasi utama perlunya dilakukan studi komparasi antara algoritma *K-Means* dan *K-Medoids* (PAM) guna menguji ketahanan algoritma terhadap distorsi data dalam memetakan elemen pesan secara presisi [16].
- Karakteristik pasar lokal, sektor kuliner di Tembilahan memiliki kompetisi visual yang tinggi. Oleh karena itu, elemen taktis seperti pemanfaatan fitur interaktif dan gaya penyampaian pesan pada akun UMKM X menjadi krusial untuk dievaluasi secara makro menggunakan teknik *data mining*.
- Kriteria inklusi kelayakan akun, objek terpilih harus menunjukkan konsistensi publikasi dan memiliki metrik interaksi publik (*likes* dan *comments*) yang valid untuk dianalisis berdasarkan standar audit efektivitas konten media [10]. Akun resmi UMKM X memenuhi prasyarat tersebut, di mana data historis selama periode Mei 2024 – Mei 2026 terbukti menyediakan volume data yang aktif, variatif, dan representatif untuk kebutuhan ekstraksi pola berbasis *data mining*.



3.3.2 Definisi Operasional Variabel

Sebagai fase awal, dilakukan ekstraksi data secara kuantitatif deskriptif terhadap dataset historis untuk menghasilkan variabel operasional guna menemukan pola performa tersembunyi (*hidden patterns*) berdasarkan kerangka Ambarsari dkk. [7]. Untuk kebutuhan komputasi, variabel kualitatif (pesan dan fitur interaktif) dikonversi ke dalam bentuk data biner (0 atau 1) guna mendampingi nilai numerik *Engagement Rate* sebagai fitur utama pengelompokkan klaster. Sementara itu, variabel format visual dan waktu publikasi dikonversi melalui pengodean label (*label encoding*). Klasifikasi seluruh variabel dan hubungannya dengan model AIDA dirinci pada Tabel 3.1 berikut.

Tabel 3.1 Operasional Variabel Penelitian Berbasis Model AIDA

No	Nama Variabel	Konsep Teoritis (AIDA)	Indikator Variabel & Skala Data	Peran Dalam K-Means & K-Medoids	Justifikasi Literatur
1	Bahasa Promosi	<i>Desire</i>	Penggunaan gaya bahasa persuasif (emosional atau personalisasi). Skala biner: 0 = Tidak Ada, 1 = Ada.	Variabel Karakteristik (Kontekstual Klaster)	Diadopsi dari aspek emosi dan personalisasi pesan model AIDA [12].
2	Call To Action (CTA)	<i>Action</i>	Instruksi langsung untuk mendorong respons (misal: "Klik", "Beli"). Skala biner: 0 = Tidak Ada, 1 = Ada.	Variabel Karakteristik (Kontekstual Klaster)	Diadopsi dari elemen instruksi tindakan langsung pada caption [12].
3	Akses Transaksi	<i>Action</i>	Kehadiran fitur interaktif berupa tautan pembelian atau kontak pada caption atau bio. Skala biner: 0 = Tidak Ada, 1 = Ada.	Variabel Karakteristik (Kontekstual Klaster)	Diadopsi dari kemudahan konversi tahap <i>action</i> [13] [14].
4	Business Tag	<i>Action</i>	Penyebutan (mention/tag) akun bisnis lain atau mitra kolaborasi pada unggahan. Skala biner: 0 = Tidak Ada, 1 = Ada.	Variabel Karakteristik (Kontekstual Klaster)	Diadopsi dari kemudahan konversi tahap <i>action</i> [13].



Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial.
2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan.
3. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi.
4. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
5. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Universitas Islam Indragiri

Tabel 3.1 Operasional Variabel Penelitian Berbasis Model AIDA (Lanjutan)

No	Nama Variabel	Konsep Teoritis (AIDA)	Indikator Variabel & Skala Data	Peran Dalam <i>K-Means</i> & <i>K-Medoids</i>	Justifikasi Literatur
5	Engagement Rate (ER)	Action Evaluation	Rasio interaksi publik (ER): $\frac{\text{Jumlah Likes} + \text{Komentar}}{\text{Total Pengikut}} \times 100\%$ Skala: Numerik (rasio)	Variabel Fitur Utama & Evaluasi	Diadopsi dari metrik evaluasi performa kluster komunikasi [10][17].
6	Hari Unggah	Attention	Hari publikasi konten pada Instagram (Senin s.d Minggu) Skala: Kategorikal / Nominal	Variabel Karakteristik (Kontekstual Klaster)	Diadopsi dari aspek manajemen waktu distribusi konten [5].
7	Waktu Unggah	Attention	Waktu rilis konten berdasarkan format waktu 24 jam. Skala: Kategorikal / Interval	Variabel Karakteristik (Kontekstual Klaster)	Diadopsi dari pola waktu aktif audiens media sosial [5]
8	Tipe Konten	Attention & Interest	Format visual publikasi unggahan pada akun Instagram. Skala Kategorikal (Koding): 1 = <i>Single Photo / Image</i> 2 = <i>Carousel</i> 3 = <i>Reels / Video</i>	Variabel Karakteristik (Kontekstual Klaster)	Diadopsi berdasarkan preferensi format visual yang efektif dalam menangkap perhatian awal audiens [6].

Berdasarkan Tabel 3.1, variabel penelitian disusun mengacu pada hierarki model AIDA. Waktu dan tipe konten mewakili tahap *Attention* dan *Interest*, Bahasa Promosi mewakili *Desire*, serta CTA, *Business Tag*, dan Akses Transaksi mewakili *Action*. Selanjutnya, metrik *Engagement Rate* (ER) digunakan sebagai evaluasi keterlibatan riil audiens (*Action Evaluation*) [10, 16].

Guna menghindari bias skala matematis saat komputasi klasterisasi (*K-Means* dan *K-Medoids*), variabel dibagi menjadi dua peran:

1. Fitur Utama (*Clustering Feature*): Hanya menggunakan nilai ER sebagai variabel input tunggal algoritma untuk membentuk kluster performa konten.



2. Fitur Karakteristik (*Contextual Features*): Meliputi elemen AIDA biner, konteks waktu, dan tipe konten yang difungsikan sebagai variabel meta pascaklasterisasi untuk memprofilkan karakteristik strategi pada setiap klaster.

3.2 Perangkat Pendukung Penelitian

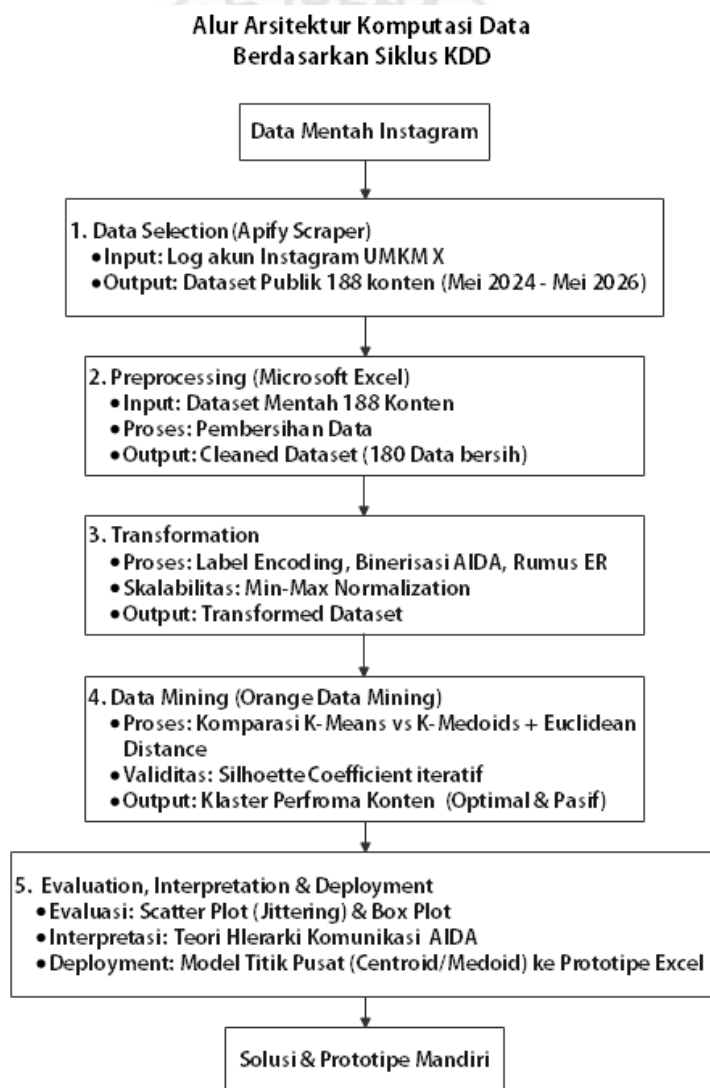
Instrumen dan alat penelitian dirancang untuk mendukung alur kerja KDD yang terbagi ke dalam dua kategori:

1. Instrumen Penelitian, berupa Lembar Kodifikasi Data (*Data Coding Sheet*) berbasis Microsoft Excel untuk mencatat metrik numerik dan kontekstual (*likes*, komentar, waktu, format) mewakili *Attention* dan *Interest*, serta pengodean variabel biner (0 atau 1) untuk keberadaan Bahasa Promosi (*Desire*), CTA, *Business Tag*, dan Akses Transaksi (*Action*).
2. Alat Penelitian (Tools), menggunakan satu unit laptop spesifikasi standar serta didukung beberapa perangkat lunak, yaitu:
 - a) Apify (Instagram Scraper): Mengekstraksi 188 data unggahan publik UMKM X periode dua tahun.
 - b) Microsoft Excel: Pra-pemrosesan, tabulasi data, pengodean biner, dan pembuatan prototipe evaluasi.
 - c) Orange Data Mining: Pemodelan klasterisasi (K-Means vs. K-Medoids) dan uji Silhouette Coefficient.
 - d) Aplikasi Instagram, Browser, & MS Word: Untuk validasi silang data riil serta penyusunan dokumen laporan.



3.3 Prosedur Penelitian: Kerangka Kerja KDD

Prosedur penelitian ini mengikuti tahapan *Knowledge Discovery in Databases* (KDD) yang diadaptasi dari metodologi [5]. Alur komputasi ini dirancang untuk mengekstraksi pola informasi strategis dari data log akun Instagram resmi UMKM X melalui lima tahapan utama yang saling berkesinambungan. Untuk menggambarkan bagaimana data mentah ditransformasikan hingga menjadi model pengetahuan digital, arsitektur aliran data dalam siklus KDD dirinci pada Gambar 3.3 berikut.



Gambar 3.3 Alur Arsitektur Komputasi Data Berdasarkan Siklus KDD





Gambar 3.3 menyajikan arsitektur aliran data yang mengilustrasikan transformasi dari data mentah hingga menjadi luaran strategis berbasis siklus *Knowledge Discovery in Databases* (KDD). Secara hierarkis, alur komputasi ini terbagi menjadi lima fase utama:

1. Pemilihan Data (*Data Selection*)

Pengambilan 188 data log publik historis akun UMKM X periode Mei 2024 – Mei 2026 menggunakan *Apify Instagram Scraper* setelah dilakukan validasi indikator pada *pilot study*.

2. Pemrosesan Awal (*Preprocessing*)

Pembersihan data dari *missing values* serta eliminasi unggahan non-bisnis berbasis Microsoft Excel hingga menghasilkan 180 dataset bersih.

3. Transformasi Data (*Transformation*)

Tahap transformasi mengubah data mentah kualitatif dan kuantitatif ke dalam struktur matematika yang siap diproses algoritma *clustering*, meliputi:

- 1) Pengodean Kategori (*Label Encoding*): Mengonversi data waktu dan format visual menjadi indeks angka, yaitu hari unggah (0–6), jam unggah (0–23), serta tipe konten (1 = single photo, 2 = carousel, 3 = reels/video).
- 2) Kodifikasi Data Biner: Mengubah karakteristik pesan kualitatif pada *caption* menjadi nilai biner (0 = tidak ada, 1 = ada) untuk variabel bahasa promosi, CTA, akses transaksi, dan *business tag*.
- 3) Kalkulasi metrik, Menghitung nilai *Engagement Rate* (ER) berbasis metrik *likes* dan *comments* menggunakan Persamaan 3.1:

$$\text{Engagement Rate (ER)} = \frac{\text{Likes} + \text{Comments}}{\text{Total Followers}} \times 100\%$$

- 4) Normalisasi data, Menerapkan *min-max normalization* pada fitur numerik untuk menyetarakan skala variabel dan mencegah bias jarak absolut:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Keterangan:

- a) X_{norm} = Nilai data setelah dinormalisasi
- b) X = Nilai data asli sebelum dinormalisasi
- c) X_{min} = Nilai minimum dari variabel tersebut dalam seluruh dataset
- d) X_{max} = Nilai maksimum dari variabel tersebut dalam seluruh dataset.

4. Penambangan Data (*Data Mining*)

Dataset hasil transformasi diimpor ke *software* Orange Data Mining untuk membandingkan kinerja algoritma *K-Means* dan *K-Medoids* (PAM). Jarak antarobjek dihitung menggunakan *Euclidean Distance* (Persamaan 3.3):

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_{ik})^2}$$

Keterangan:

- $d(x, y)$ = jarak *Euclidean* antara objek data x dan titik pusat y .
- X_k = Nilai ternormalisasi dari variabel fitur ke- k pada data yang diukur.
- Y_{ik} = Nilai titik pusat kluster ke- i pada variabel fitur ke- k . Titik pusat berupa nilai rata-rata (*Centroid*, C_i) pada algoritma *K-Means*, atau objek rill (*Medoid*, M_i) pada algoritma *K-Medoids*.
- n = Jumlah total variabel fitur (*Clustering Features*).

Penentuan jumlah kluster (K) optimal diuji secara objektif menggunakan

Silhouette Coefficient (Persamaan 3.4):

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Keterangan:

- a) $s(i)$ = Skor *Silhouette Coefficient* objek ke- i .
- b) $a(i)$ = Rata-rata jarak objek ke- i di dalam klasternya sendiri.
- c) $b(i)$ = Rata-rata jarak objek ke- i dengan kluster tetangga terdekat.



Untuk memvalidasi bahwa perbedaan karakteristik antar-klastr signifikan secara statistik, uji *Analysis of Variance* (ANOVA) diintegrasikan setelah nilai K optimal diperoleh [17]. Komputasi dihentikan saat mencapai tahap konvergensi, yaitu ketika posisi *Centroid* (*K-Means*) atau titik *Medoid* (*K-Medoids*) telah stabil dan tidak mengalami perubahan.

5. Evaluasi dan Interpretasi (*Evaluation, Interpretation, and Deployment*)

Tahap akhir ini menerjemahkan hasil klasterisasi menjadi rekomendasi manajerial aplikatif melalui tiga aktivitas:

- 1) Evaluasi dan visualisasi, menilai pemisahan klaster dan signifikansi ANOVA menggunakan tabel ringkasan, *box plot*, dan *scatter plot* berfitur *jittering*.
- 2) Interpretasi teoritis, membedah pola variabel waktu dan tipe konten berbasis hierarki model AIDA pada klaster hasil algoritma unggulan.
- 3) Penerapan (*deployment*), mentransformasikan titik pusat akhir (*Centroid/Medoid*) algoritma terbaik ke fungsi logika operasional lembar kerja Microsoft Excel [17]. Label kelompok data baru ditentukan berdasarkan jarak *Euclidean* terkecil terhadap titik pusat klaster:

$$\text{Status Klaster} = \arg \min_{k \in \{1,2,\dots,K\}} (\text{Jarak}_{Ck})$$

Keterangan:

- a) Jarak_{Ck} = Nilai kedekatan (jarak *Euclidean*) objek data baru terhadap titik pusat (*Centroid* atau *Medoid*) klaster ke- k .
- b) K = Jumlah total klaster optimal yang terbentuk.

Instrumen teknis interaktif ini memungkinkan pemilik akun UMKM

X mengevaluasi kinerja konten baru secara mandiri dan berkelanjutan.

