

CEK TURNITIN

SKRIPSI Febriyola Desy Azzara

 SATSET DIGITAL 1

Document Details

Submission ID

trn:oid:::3618:151242494

Submission Date

Sep 10, 2026, 2:40 PM GMT+7

Download Date

Sep 10, 2026, 2:49 PM GMT+7

File Name

SKRIPSI Febriyola Desy Azzara.pdf

File Size

1.4 MB

64 Pages

10,496 Words

71,048 Characters

22% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Top Sources

- 11%  Internet sources
- 15%  Publications
- 13%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

- 11% Internet sources
- 15% Publications
- 13% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Publication	Febriyola Desy Azzahra, Fitri Yunita, Bayu Rianto. "Analisis Komparatif Karakteris...	7%
2	Internet	repository.unisi.ac.id	2%
3	Internet	dspace.uil.ac.id	<1%
4	Internet	kc.umh.ac.id	<1%
5	Internet	digilib.uinsgd.ac.id	<1%
6	Student papers	Academic Library Consortium on 2025-02-26	<1%
7	Internet	repository.uin-suska.ac.id	<1%
8	Student papers	Universitas Tarumanagara on 2025-12-17	<1%
9	Internet	journal.ilmudata.co.id	<1%
10	Student papers	Universitas Negeri Jakarta on 2017-07-11	<1%
11	Student papers	Universitas Islam Riau on 2026-05-21	<1%

12	Student papers	Universitas Islam Riau on 2026-06-04	<1%
13	Publication	Nur Rohman, Arief Wibowo. "Perbandingan Metode K-Medoids dan Metode K-Me..."	<1%
14	Student papers	Tarumanagara University on 2026-06-21	<1%
15	Student papers	Universitas Tarumanagara on 2025-12-27	<1%
16	Student papers	Tarumanagara University on 2026-06-21	<1%
17	Internet	repository.wima.ac.id	<1%
18	Publication	Akbar Suseno Tri Maulana, Achmad Lutfi Fuadi. "PERBANDINGAN METODE HASIL ..."	<1%
19	Student papers	Universitas Islam Indonesia on 2018-08-10	<1%
20	Internet	digilib.ulm.ac.id	<1%
21	Internet	journal.ummat.ac.id	<1%
22	Internet	jurnal.polibatam.ac.id	<1%
23	Internet	repository.peradaban.ac.id	<1%
24	Internet	123dok.com	<1%
25	Internet	digilib.isi.ac.id	<1%

26	Internet	repository.its.ac.id	<1%
27	Publication	M Datar Tafakur Dharma Putra, Ainur Rocmaniah. "Menganalisis Pesan Perawata...	<1%
28	Internet	journal.foundae.com	<1%
29	Publication	Hillary Octa Fatica, Aviv Fitria Yulia, Ferly Ardhy, Fahlul Rizki. "Perbandingan Kine...	<1%
30	Student papers	Sekolah Teknik Elektro & Informatika on 2025-07-18	<1%
31	Student papers	UPN Veteran Yogyakarta on 2025-10-28	<1%
32	Student papers	Universitas Islam Riau on 2025-07-17	<1%
33	Internet	begawe.unram.ac.id	<1%
34	Internet	lab.tik.pnj.ac.id	<1%
35	Internet	melekit-if.uwks.ac.id	<1%
36	Internet	pmc.ncbi.nlm.nih.gov	<1%
37	Internet	repository.ittelkom-pwt.ac.id	<1%
38	Internet	repository.pnj.ac.id	<1%
39	Internet	repository.ub.ac.id	<1%

40	Student papers	Universitas Tarumanagara on 2025-12-24	<1%
41	Student papers	Universitas Brawijaya on 2017-06-15	<1%
42	Internet	repository.nusaputra.ac.id	<1%
43	Student papers	Tarumanagara University on 2026-06-21	<1%
44	Student papers	UPN Veteran Yogyakarta on 2026-08-26	<1%
45	Student papers	Universitas Dian Nuswantoro on 2016-02-10	<1%
46	Student papers	Universitas Tarumanagara on 2025-12-18	<1%
47	Internet	eprints.undip.ac.id	<1%
48	Internet	id.scribd.com	<1%
49	Internet	jusinfo.upjb.ac.id	<1%
50	Internet	repository.ppns.ac.id	<1%
51	Internet	repository.uph.edu	<1%
52	Internet	seniurutterapisifuli.wordpress.com	<1%
53	Internet	www.adi-journal.org	<1%

54	Internet	www.mdpi.com	<1%
55	Student papers	Catholic University of Parahyangan on 2023-07-09	<1%
56	Publication	Dennis Ma'rifatul Kurnia, Nurun Nihayatur Rifqiya Aulia, Yulistiya Nur Hidayah, Ci...	<1%
57	Student papers	LSPR Communication & Business Institute on 2025-05-08	<1%
58	Publication	Penda Sudarto Hasugian, Agustina Simangunsong, Meriana Manik, Hotnitaria T...	<1%
59	Student papers	Tarumanagara University on 2026-06-21	<1%
60	Student papers	Universitas Brawijaya on 2017-02-23	<1%
61	Student papers	Universitas Putera Batam on 2026-07-31	<1%
62	Internet	ejurnal.seminar-id.com	<1%
63	Internet	id.123dok.com	<1%
64	Internet	juris.id	<1%
65	Internet	repository.poltekeskupang.ac.id	<1%
66	Internet	text-id.123dok.com	<1%
67	Internet	www.docstoc.com	<1%

68	Publication	Ante Wahyu Alvianingrum, Annisa Noor Rachmawati. "Pengembangan Strategi B...	<1%
69	Publication	I Putu Arya Vidyananta Vidyananta, Kadek Teguh Dermawan Dermawan. "PERBA...	<1%
70	Publication	Novianto puji Raharjo. "Digital Identity and Ethical Challenges on Social Media: A...	<1%
71	Publication	Rizma Nurhaliza, Windarsyah Windarsyah, Finki Dona Marleny. "KLASTERISASI ST...	<1%
72	Student papers	Universitas 17 Agustus 1945 Semarang on 2026-09-01	<1%
73	Student papers	Universitas Brawijaya on 2018-01-03	<1%
74	Student papers	Universitas Pendidikan Indonesia on 2014-12-17	<1%
75	Internet	digilib.ptdisttd.ac.id	<1%
76	Internet	docplayer.info	<1%
77	Internet	repositori.umrah.ac.id	<1%
78	Internet	repositori.unimma.ac.id	<1%
79	Internet	repository.ummy.ac.id	<1%
80	Internet	web.unikom.ac.id	<1%
81	Student papers	East Union High School on 2026-07-02	<1%

82	Publication	Maria Titah Jatipaningrum, Suci Eka Azhari, Kris Suryowati. "Pengelompokan Kab...	<1%
83	Student papers	Tarumanagara University on 2024-12-02	<1%
84	Student papers	Universitas Mulawarman on 2021-01-11	<1%
85	Student papers	Universitas Tarumanagara on 2025-12-30	<1%
86	Student papers	Sriwijaya University on 2019-12-20	<1%
87	Student papers	The University of the West of Scotland on 2025-06-30	<1%
88	Student papers	UIN Sultan Syarif Kasim Riau on 2021-08-16	<1%
89	Student papers	UPN Veteran Yogyakarta on 2025-01-31	<1%
90	Student papers	Universitas Brawijaya on 2018-10-23	<1%

43

**ANALISIS KOMPARATIF ALGORITMA K-MEANS DAN K-MEDOIDS
DALAM PENGELOMPOKKAN ENGAGEMENT KONTEN PEMASARAN**

PROPOSAL



Disusun Oleh:

FEBRIYOLA DESY AZZAHRA

403221010064

2

PRODI SISTEM INFORMASI

FAKULTAS TEKNIK DAN ILMU KOMPUTER

UNIVERSITAS ISLAM INDRAGIRI

TAHUN 2025/2026

KATA PENGANTAR

Karya tulis ini disusun sebagai hasil penelitian untuk syarat mendapatkan gelar sarjana S1. Dalam menyelesaikan Tugas Akhir, penulis banyak mendapatkan bimbingan, bantuan dan dukungan yang sangat berarti dari berbagai pihak sehingga proposal ini dapat tersusun dengan baik. Untuk itu penulis mengucapkan banyak terima kasih kepada:

1. Bapak Dr. H. Najamuddin, Lc., M. A, selaku Rektor Universitas Islam Indragiri, yang telah memfasilitasi, mendukung, dan memberikan kesempatan berharga kepada penulis, sehingga proses penyusunan proposal dapat berjalan dengan baik.
2. Bapak. Dr. Bayu Rianto, S.SI, M. Kom., selaku Dekan Fakultas Teknik dan Ilmu Komputer sekaligus Dosen Pembimbing pendamping, yang dengan penuh dedikasi telah memberikan bimbingan, masukan, pengetahuan, serta arahan yang sangat berharga sehingga proposal ini dapat diselesaikan dengan baik.
3. Ibu Fitri Yunita S. SI., M. Kom., selaku Ketua Program Studi Sistem Informasi sekaligus Dosen Pembimbing utama, yang telah memberikan bimbingan, perhatian, dukungan, serta masukan yang berarti selama proses penyusunan proposal hingga terselesaikan dengan baik.
4. Bapak Abdul Muni, S. Kom., M. Kom., selaku Dosen Pembimbing Akademik, yang telah memberikan bimbingan, arahan, serta dukungan akademik selama menjalani proses perkuliahan hingga penyusunan proposal ini.

- 2 5. Seluruh dosen Sistem Informasi dan staf akademik Universitas Islam Indragiri, yang telah memberikan ilmu, pelayanan, dan dukungan selama penulis menempuh pendidikan.
- 20
- 51 6. Kedua orang tua tercinta, yang selalu menjadi sumber semangat, doa, dan dukungan tanpa henti kepada penulis dalam menyelesaikan setiap tahap studi dan penyusunan proposal ini.
- 47
- 2 7. Teman-teman seperjuangan, sahabat, dan rekan diskusi yang senantiasa memberikan dukungan, motivasi, serta membantu penulis dalam menghadapi proses penyusunan proposal ini.

Dalam hal ini penulis menyadari sepenuhnya bahwa dengan bekal ilmu dan kemampuan yang terbatas, tidaklah mudah untuk membuat suatu laporan penelitian yang sempurna, oleh sebab itu dengan segala kerendahan hati dan rasa terima kasih yang dalam, penulis mengharapkan saran-saran dan kritik yang membangun demi kesempurnaan penelitian ini.

Akhir kata penulis berharap semoga laporan ini dapat memberikan manfaat bagi rekan-rekan Universitas Islam Indragiri dan para pembaca.

Tembilahan, 30 Oktober 2025

FEBRIYOLA DESY AZZAHRA
403221010064

DAFTAR ISI

KATA PENGANTAR.....	i
DAFTAR ISI.....	iv
DAFTAR GAMBAR	v
DAFTAR TABEL	vi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Penelitian	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Penelitian.....	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	8
2.1 Kajian Penelitian Terdahulu	8
2.2 Rangkuman	15
2.3 Kesimpulan dan Kebaruan Penelitian	20
BAB III METODE PENELITIAN	24
3.1 Desain Penelitian dan Definisi Operasional Variabel	24
3.2 Perangkat Pendukung Penelitian.....	32
3.3 Prosedur Penelitian: Kerangka Kerja KDD.....	33
BAB IV HASIL DAN PEMBAHASAN	37
4.1 Hasil Pengolahan Data.....	37
4.2 Hasil Transformasi	39
4.3 Hasil Implementasi	41
4.4 Hasil Evaluasi dan Kontekstualisasi Temuan Model AIDA	50
4.5 Implikasi Operasional dan Pemeliharaan Model.....	51
4.6 Rangkuman	52
BAB V PENUTUP	54
5.1 Kesimpulan.....	54
5.2 Saran.....	54
DAFTAR PUSTAKA.....	55
LAMPIRAN.....	57

DAFTAR GAMBAR

Gambar 3.1 Diagram Alir Keseluruhan Penelitian	26
Gambar 3.2 Data 20 Unggahan Study Pilot	28
Gambar 3.3 Alur Arsitektur Komputasi Data Berdasarkan Siklus KDD	33
Gambar 4.1 Potongan Lembar Kerja Dataset 2 Tahun (Mei 2024- Mei 2026).....	39
Gambar 4.2 Potongan Dataset Hasil Transformasi Matriks Numerik.....	41
Gambar 4.3 Visualisasi Nilai <i>Silhouette Coefficient</i> pada Widget <i>K-Means</i>	43
Gambar 4.4 Visualisasi Klaster Melalui Grafik Scatter Plot (bisnisTag).....	Error!
Bookmark not defined.	
Gambar 4. 5 Visualisasi Klaster Melalui Grafik Scatter Plot (CTA)	45
Gambar 4. 6 Visualisasi Klaster Melalui Grafik Scatter Plot (PromoLang)	46
Gambar 4. 7 Visualisasi Klaster Melalui Grafik Scatter Plot (AksesTransaksi)...	46
Gambar 4.8 Visualisasi Klaster Melalui Grafik Scatter Plot (TipeKonten)	46
Gambar 4.9 Box Plot ER_Normalisasi dan Klaster	48
Gambar 4.10 Distribusi Proporsi Tipe Konten Antar Klaster	48
Gambar 4. 11 Tampilan Prototipe Lembar Kerja Pemeliharaan Konten dan Simulasi Otomatisasi Klaster	52

2

DAFTAR TABEL

Tabel 2.1 Perbandingan dengan Penelitian Terdahulu	8
Tabel 2.1 Perbandingan dengan Penelitian Terdahulu (Lanjutan)	9
Tabel 2.1 Perbandingan dengan Penelitian Terdahulu (Lanjutan)	11
Tabel 2.1 Perbandingan dengan Penelitian Terdahulu (Lanjutan)	12
Tabel 2.1 Perbandingan dengan Penelitian Terdahulu (Lanjutan)	13
Tabel 2.2 Perbandingan Research Gap.....	21
Tabel 3.1 Operasional Variabel Penelitian Berbasis Model AIDA	29
Tabel 4.1 Distribusi Fekuensi Unggahan UMKM X Tahun 2024.....	37
Tabel 4.2 Distribusi Frekuensi Unggahan UMKM X Tahun 2025	38
Tabel 4.3 Distribusi Frekuensi Unggahan UMKM X Tahun 2026	38
Tabel 4.4 Aturan Pengodean (<i>Encoding</i>) Transformasi Variabel	40
Tabel 4.5 Sampel Pengelompokkan Data Mnegggunakan Algoritma K-Means.....	44

24

45

BAB I PENDAHULUAN

1.1 Latar Belakang Penelitian

Pada era digital, sektor *Food and Beverage* (F&B) sangat bergantung pada visualisasi digital untuk menarik minat konsumen. Instagram menjadi media utama pelaku usaha kuliner untuk menekan biaya pemasaran (*marketing cost*) sekaligus menjangkau target pasar secara lebih presisi. Penelitian Nasution [1] menunjukkan mayoritas konsumen memperoleh informasi dan membangun minat berkunjung melalui aktivitas di *platform* ini. Melalui media sosial, perusahaan dapat membangun identitas produk dan menciptakan keterlibatan pelanggan (*customer engagement*) yang mendalam, sebagai infrastruktur vital dalam menjaga keberlangsungan bisnis di tengah kompetisi yang ketat.

Namun, banyak pelaku usaha cenderung terjebak pada strategi *hard-selling* yang agresif, yang sering kali berbanding terbalik dengan tingkat interaksi audiens. Penelitian Salsabila [2] menegaskan bahwa pendekatan *soft-selling* justru lebih efektif dalam membangun *emotional engagement* untuk mendorong niat beli secara alami. Hal ini sejalan dengan temuan Yanti [3], bahwa strategi komunikasi pemasaran yang tidak mencapai standar rata-rata *engagement* (di bawah 5,3%) berisiko diabaikan audiens. Mengacu model AIDA (*Attention, Interest, Desire, Action*), metrik interaksi (*likes* dan *comments*) merupakan pembuktian empiris tercapainya tahap *Attention* dan *Interest*. Tanpa keseimbangan antara promosi (*Action*) dan konten keterlibatan emosional, pesan pemasaran berisiko gagal mencapai efisiensi komunikasi yang optimal.

Fenomena tersebut teramati nyata pada akun Instagram UMKM X yang bergerak sebagai media informasi publik regional dengan fokus promosi kuliner

1 lokal. Observasi awal terhadap sampel 20 postingan terbaru menunjukkan bahwa 19 di antaranya merupakan konten berformat Carousel dengan strategi hard-selling yang dominan, memenuhi indikator Business Tag, Promo Language, (Call to Action) CTA, dan Transaction Access. Selain itu, ditemukan ketidakkonsistenan distribusi waktu unggah, di mana aktivitas meningkat tajam pada periode tertentu seperti bulan Ramadan. Dominasi promosi masif yang tidak konsisten ini menghasilkan respons audiens yang rendah (rata-rata 45 Likes), mengindikasikan adanya ketidakefisienan dalam strategi komunikasi, penentuan pilar konten, dan manajemen waktu publikasi yang dijalankan.

1 Untuk mengatasi masalah ketidakefisienan tersebut, diperlukan evaluasi berbasis pada data objektif (*data-driven insights*) mencakup data historis dua tahun terakhir. Namun, pengolahan data berskala besar dan tidak terstruktur tidak lagi relevan jika dilakukan dengan rekapitulasi manual. Sebagaimana ditegaskan oleh Rahman [4] pemanfaatan data analitis merupakan strategi esensial untuk mengambil keputusan yang responsif terhadap kebutuhan konsumen. Melalui pemanfaatan perangkat lunak otomatisasi pengumpulan data (*web scraping*) dan teknik *data mining*, tumpukan data interaksi publik yang acak dapat ditransformasikan menjadi pola informasi strategis yang bernilai tinggi bagi manajemen konten pelaku usaha.

Namun demikian, penerapan algoritma *K-Means* memiliki keterbatasan mendasar ketika diimplementasikan pada data *engagement* media sosial publik. Data interaksi publik sering kali memiliki distribusi yang tidak merata dan rentan terhadap keberadaan data pencilan (*outliers*), seperti konten yang mendadak *viral* dengan angka *likes* atau *comments* yang jauh melampaui rata-rata. *K-Means*

14 sangat sensitif terhadap *outlier* karena menggunakan perhitungan titik pusat berbasis rata-rata (*mean*), sehingga berisiko menghasilkan klasterisasi yang terdistorsi. Untuk mengatasi kelemahan tersebut, penelitian ini menyandingkannya dengan algoritma *K-Medoids (Partitioning Around Medoids)*. *K-Medoids* menggunakan titik pusat berupa objek nyata (*medoids/median*), sehingga lebih resisten terhadap pencilan dan variansi tinggi pada data pemasaran digital. 12 Komparasi kedua algoritma ini menjadi krusial untuk membuktikan secara empiris metode mana yang menghasilkan pengelompokan *engagement* paling presisi dan akurat.

71 Berdasarkan permasalahan di atas, penelitian ini mengusulkan sebuah pendekatan terstruktur menggunakan kerangka kerja *Knowledge Discovery in Databases (KDD)*. Pengumpulan data historis akun selama dua tahun terakhir akan dilakukan secara otomatis menggunakan *Apify Instagram Scraper*, dilanjutkan dengan tahap transformasi dan pelabelan kategori konten berbasis model AIDA. 59 Selanjutnya, dilakukan analisis komparatif antara algoritma *K-Means* dan *K-Medoids* memanfaatkan perangkat lunak *Orange Data Mining* untuk mengelompokkan konten berdasarkan metrik interaksinya (*engagement metrics*). Performa kedua algoritma akan dievaluasi dan dibandingkan tingkat akurasi menggunakan *Silhouette Coefficient* untuk menentukan algoritma clustering yang paling optimal. Sebagai luaran aplikatif, rumusan klasterisasi terbaik yang dihasilkan akan ditransformasikan menjadi dokumen cetak biru (*blueprint*) dan panduan rekomendasi tata kelola konten. Output tersebut dirancang untuk mempermudah pemilik akun atau pelaku UMKM kuliner dalam

menyusun komposisi strategi pemasaran konten serta manajemen waktu publikasi yang lebih efisien dan tepat sasaran di masa depan berdasarkan tren data historis.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana karakteristik pola komunikasi pemasaran Instagram UMKM X ditinjau dari model AIDA dan struktur konten *hard-selling*?
2. Bagaimana perbandingan performa dan tingkat akurasi antara algoritma *K-Means* dan *K-Medoids* dalam mengelompokkan unggahan berdasarkan metrik interaksi (*engagement*) menggunakan evaluasi *Silhouette Coefficient*?
3. Bagaimana evaluasi efisiensi strategi komunikasi konten saat ini dalam memicu interaksi audiens berdasarkan karakteristik tiap klaster yang terbentuk pada algoritma terbaik hasil komparasi?

1.3 Batasan Penelitian

Agar penelitian lebih terfokus dan terarah, batasan masalah yang ditetapkan adalah:

1. Data penelitian terbatas pada informasi publik akun Instagram UMKM X selama dua tahun terakhir tanpa melibatkan metrik privat seperti *reach*, *impressions*, atau *saves*.
2. Evaluasi performa berfokus pada metrik interaksi *likes* dan *comments* (tanpa mengukur konversi penjualan riil), yang dianalisis menggunakan komparasi algoritma *K-Means* dan *K-Medoids* berbasis metrik *Silhouette Coefficient*.

3. Analisis kategori konten dibatasi pada aspek tekstual dan visual menggunakan indikator AIDA serta karakteristik konten promosi (*hard-selling*), tanpa memverifikasi status kontrak finansial di balik unggahan tersebut.

1.4 Tujuan Penelitian

Sejalan dengan rumusan masalah, tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengidentifikasi karakteristik pola komunikasi pemasaran pada akun Instagram UMKM X ditinjau dari elemen model AIDA dan struktur konten *hard-selling*.
2. Menganalisis dan membandingkan performa serta tingkat akurasi antara algoritma *K-Means* dan *K-Medoids* dalam mengelompokkan unggahan berdasarkan metrik interaksi (*engagement*) menggunakan evaluasi *Silhouette Coefficient*.
3. Mengevaluasi efisiensi strategi komunikasi konten saat ini dalam memicu interaksi audiens berdasarkan karakteristik tiap klaster yang terbentuk pada algoritma terbaik hasil komparasi, guna merumuskan panduan rekomendasi tata kelola konten yang tepat sasaran.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi sebagai berikut:

1. Menjadi referensi akademis mengenai komparasi efektivitas *K-Means* dan *K-Medoids* dalam mengelompokkan metrik *engagement* media sosial yang rentan *outliers*.

2. Memberikan masukan konkret bagi UMKM X dalam mengevaluasi dan menyusun tata kelola konten berdasarkan karakteristik klaster terbaik.
3. Memberikan gambaran praktis pemanfaatan *data mining* dan kerangka AIDA untuk mendukung pemasaran berbasis data (*data-driven marketing*) pada UMKM.

1.6 Sistematika Penulisan

Berikut merupakan sistematika penulisan penelitian ini:

1. BAB 1 : PENDAHULUAN

Menguraikan latar belakang mengenai efisiensi komunikasi konten dan urgensi komparasi algoritma, rumusan masalah, batasan penelitian, tujuan, manfaat, serta sistematika penulisan.

2. BAB 2: TINJAUAN PUSTAKA

Menyajikan landasan teori meliputi pemasaran media sosial, model AIDA, strategi konten *hard-selling*, algoritma *K-Means*, *K-Medoids*, serta metrik evaluasi *Silhouette Coefficient*. Bab ini juga memuat telaah 10 penelitian terdahulu untuk memetakan posisi dan kebaruan (*novelty*) riset.

3. BAB 3: METODE PENELITIAN

Menjelaskan alur penelitian berbasis kerangka kerja *Knowledge Discovery in Databases* (KDD), mencakup ekstraksi data publik Instagram, *preprocessing*, transformasi/kodifikasi elemen AIDA, serta prosedur perancangan model komparasi *K-Means* dan *K-Medoids*.

4. BAB 4: HASIL DAN PEMBAHASAN

13

Menyajikan hasil olah data klasterisasi *K-Means* dan *K-Medoids*, evaluasi komparasi tingkat akurasi menggunakan *Silhouette Coefficient*, serta analisis karakteristik konten AIDA berdasarkan klaster terbaik yang terbentuk.

5. BAB 5: KESIMPULAN DAN SARAN

Menyajikan kesimpulan utama terkait algoritma clustering yang paling optimal beserta temuan profil kontennya, serta memberikan saran praktis bagi UMKM X dan saran akademis untuk penelitian selanjutnya.

44

44

BAB II TINJAUAN PUSTAKA

6

Bab ini menguraikan telaah terhadap minimal 10 artikel jurnal menggunakan analisis SWOT guna memetakan posisi, persamaan, dan kebaruan (*novelty*) riset pada objek penelitian. Bagian akhir bab menyajikan sintesis integrasi teori pemasaran media sosial, model AIDA, strategi *hard-selling*, serta landasan teori dan rumus matematis algoritma *K-Means*, *K-Medoids*, dan evaluasi *Silhouette Coefficient* dalam kerangka KDD yang diadopsi.

2.1 Kajian Penelitian Terdahulu

32

Sub-bab ini menelaah minimal 10 artikel jurnal relevan untuk membandingkan metodologi, variabel, dan objek penelitian guna memperkuat landasan teoretis serta mengidentifikasi celah penelitian (*research gap*), khususnya terkait komparasi performa algoritma *clustering*. Detail komparasi metode, variabel, dan temuan utama disajikan dalam Tabel 2.1, dilanjutkan dengan rangkuman kontribusi spesifik bagi studi ini.

9

9

Tabel 2.1 Perbandingan dengan Penelitian Terdahulu

No.	Nama Peneliti	Judul Penelitian	Pembahasan
(1)	(2)	(3)	(4)
1.	Lubis & Iqbal (2026) [5]	ANALISIS WAKTU OPTIMAL UNGGAH INSTAGRAM MENGGUNAKAN <i>K-MEANS CLUSTERING</i> PADA DATA SINTETIS	- Metode: <i>Knowledge Discovery In Databases</i> (KDD), algoritma <i>K-Means Clustering</i> dan <i>Min-Max Normalization</i> pada data sintesis. - Variabel: Hari unggah (kode 0-6), Jam unggah (kode 0-23), dan metrik Interaksi. - Hasil: Waktu unggah optimal pada hari Sabtu pukul 09:00 WIB. - Relevansi: Dasar penerapan KDD, pengodean waktu, dan normalisasi, serta mengadopsi rumus <i>Euclidean</i> dan <i>centroid</i> untuk evaluasi otomatis waktu unggah dan karakteristik konten UMKM X.

Tabel 2.2 Perbandingan dengan Penelitian Terdahulu (Lanjutan)

(1)	(2)	(3)	(4)
2.	Nurhidayat, dkk (2025) [6]	K-MEANS BASED BEHAVIORAL SEGMENTATION OF SOCIAL MEDIA USERS FOR DIGITAL COMMUNICATION ANALYSIS IN INDONESIA	- Metode: <i>K-Means Clustering</i>, <i>Elbow Method</i> dan <i>Silhouette Coefficient</i> pada 553 responden. - Varibel: Perilaku Pengguna, Preferensi Platform, Frekuensi Akses, dan Segmentasi Komunikasi. - Hasil: Segmentasi menjadi dua klaster, 0 (audiens aktif, responsif terhadap video pendek hiburan/emosional) dan klaster 1 (audiens pasif, membutuhkan konten informatif-kredibel untuk edukasi). - Relevansi: Acuan penggunaan <i>Silhouette Coefficient</i> untuk validasi jumlah klaster (K), landasan pemilihan variabel Tipe Konten (<i>Reels</i>) sebagai pemicu <i>Attention</i>, dan metode <i>Label Encoding</i> pada data kontekstual.
3.	Ambarsari, dkk (2024) [7]	UTILIZING K-MEANS CLUSTERING TO UNDERSTANDING AUDIENCE INTEREST IN SEO-OPTIMIZED MEDIA CONTENT	- Metode: <i>K-Means</i> dan <i>Elbow Method</i> pada 327 queries data log riil. - Variabel: Impresi (<i>Impressions</i>), Jumlah Klik (<i>Clicks</i>), Kueri Pencarian, dan Tingkat Ketertarikan. - Hasil: Identifikasi 4 karakteristik audiens, termasuk kelompok minat pasif (Impresi tinggi, klik rendah). - Relevansi: Validasi penggunaan data log riil media sosial dalam mengevaluasi respons audiens berdasarkan metrik interaksi publik (<i>Likes</i> dan <i>Comments</i>) pada objek UMKM X.
4.	Kedi, dkk (2024) [8]	MACHINE LEARNING SOFTWARE FOR OPTIMIZING SME SOCIAL MEDIA MARKETING CAMPAIGNS	- Metode: Tinjauan Pustaka (<i>Review Paper</i>) - Variabel: Adopsi digital, kampanye pemasaran media sosial, dan keberlanjutan bisnis UMKM. - Hasil: Adaptasi teknologi digital sangat esensial bagi keberlanjutan bisnis UMKM meski di tengah keterbatasan teknis. - Relevansi: Landasan teoritis mengenai urgensi <i>Machine Learning</i> bagi UMKM, sekaligus menjadi dasar motivasi implementasi algoritma <i>K-Means</i> pada objek UMKM X.

Tabel 2.3 Perbandingan dengan Penelitian Terdahulu (Lanjutan)

(1)	(2)	(3)	(4)
5.	Nallasivam, dkk (2025) [9]	PREDICTING CONSUMER ADOPTION OF LUXURY PRODUCTS VIA INSTAGRAM MARKETING: A MACHINE LEARNING APPROACH	• Metode: Metode campuran (<i>mixed-methods</i>) menggunakan statistik inferensial, <i>Sentimen Analysis</i>, dan algoritma <i>Machine Learning</i> (SVM dan <i>Decision Tree</i>). • Variabel: Metrik keterlibatan Instagram, jenis konten, faktor psikologis (kepercayaan, prestise, keunikan merek), demografi, dan adopsi produk. • Hasil: Keterlibatan permukaan tidak signifikan (<i>scrolling</i>, lihat iklan, dsb), sedangkan rekomendasi personal dan kepercayaan merek berbasis konten interaktif yang paling dominan memengaruhi minat beli. • Relevansi: Menekankan pentingnya fitur interaktif Instagram sebagai landasan untuk menguji pengaruh bisnis tag terhadap tindakan (<i>Action</i>) akses transaksi.
6.	Nugraha, dkk. (2025) [10]	ANALISIS ISI KONTEN MEDIA SOSIAL INSTAGRAM PEMERINTAH INDONESIA DI TINGKAT KEMENTERIAN DAN LEMBAGA	• Metode: Kuantitatif deskriptif dan analisis isi (<i>content analysis</i>) pada 100 instansi. • Variabel/Fokus Analisis: Kuantitas unggahan, substansi isi pesan (figur pemimpin vs. informasi publik), dan efektivitas interaksi. • Hasil: Kuantitas konten tinggi tidak menjamin efektivitas interaksi; konten masih didominasi figur pemimpin dibanding informasi publik. • Relevansi: Dasar argumen standarisasi konten pada klaster berperforma rendah, sekaligus acuan rumus <i>Engagement Rate</i> yang disesuaikan dengan ketersediaan data publik penelitian (hanya menggunakan metrik <i>likes</i> dan <i>comments</i>)

Tabel 2.4 Perbandingan dengan Penelitian Terdahulu (Lanjutan)

(1)	(2)	(3)	(4)
7.	Nafsyah, dkk. (2022) [11]	ANALISIS KONTEN MEDIA SOSIAL INSTAGRAM BY.U SEBAGAI MEDIA PENYEBARAN INFORMASI & KOMUNIKASI	• Metode: Kualitatif dengan analisis isi menggunakan teori Empat Pilar Strategi Media Sosial. • Variabel/Fokus Analisis: Jenis konten (kolaborasi, hiburan, edukasi, informasi) dan tingkat interaksi audiens (<i>audience engagement</i>). • Hasil: Konten kolaborasi dominan secara kuantitas, namun jenis konten hiburan menjadi yang paling efektif dalam menarik interaksi. • Relevansi: Sebagai dasar pengelompokan jenis konten (promosi usaha vs. konten komunitas) serta acuan teori untuk menganalisis bagaimana perbedaan karakteristik konten tersebut memengaruhi penempatan fitur interaktif seperti <i>business tag</i> dan tingkat <i>engagement</i> audiens.
8.	Yadav, Tripathi, dan Shukla (2025) [12]	AIDA MODEL FOR SOCIAL MEDIA MARKETING	• Metode: Analisis konten kualitatif model AIDA. • Variabel/Fokus Analisis: Elemen Hirarki AIDA (<i>Attention, Interest, Desire, Action</i>), Stimulus Visual, Emosi, dan Personalisasi Konten. • Hasil: Visual kuat memicu <i>Attention</i>, sedangkan emosi dan personalisasi mendorong <i>Interest</i> hingga <i>Action</i>. • Relevansi: Sebagai landasan teoritis untuk memetakan strategi tahap <i>Action</i> (melalui variabel CTA dan akses transaksi), menjadi dasar pembenaran penggunaan <i>Engagement Metrics</i> (metrik <i>likes</i> dan <i>comments</i>), serta acuan dalam mengkaji karakteristik bahasa promosi melalui pendekatan elemen emosional dan personalisasi konten.

Tabel 2.5 Perbandingan dengan Penelitian Terdahulu (Lanjutan)

(1)	(2)	(3)	(4)
9.	Rahardja, dkk. (2026) [13]	OPTIMIZING DIGITAL PROMOTIONAL GRAPHIC DESIGN STRATEGIES USING THE AIDA MODEL	• Metode: Kualitatif deskriptif dengan analisis visual dan metrik keterlibatan. • Variabel/Fokus Analisis: Strategi desain grafis promosional, elemen struktur visual AIDA, dan penempatan <i>Call to Action</i> (CTA). • Hasil: Kombinasi model AIDA dan analisis data metrik terbukti meningkatkan efektivitas promosi melalui perancangan visual yang terstruktur dan penempatan CTA yang strategis. • Relevansi: Landasan teoritis guna menganalisis bagaimana struktur visual konten promosi dan penempatan elemen <i>Action</i> (seperti CTA & <i>business tag</i>) memengaruhi pembentukan klaster performa interaksi audiens.
10.	Sarfraz, dkk. (2025) [14]	HOW DIGITAL MARKETING SHAPES CONSUMER DECISION-MAKING EMPLOYING (AIDA) MODEL WITH RESPECT TO CONSUMER KNOWLEDGE AND CONSUMER EXPERIENCE	• Metode: Kualitatif dengan pendekatan analisis tematik (<i>Thematic Analysis</i>) berdasarkan model AIDA. • Variabel/Fokus Analisis: Proses pengambilan keputusan konsumen (<i>consumer decision-making</i>), pengetahuan konsumen (<i>customer knowledge</i>), pengalaman konsumen (<i>customer experience</i>), dan model AIDA. • Hasil: Personalisasi konten meningkatkan minat (<i>interest</i>), sementara ketersediaan alat keterlibatan yang praktis (<i>user-friendly engagement tools</i>) secara signifikan mendorong audiens menuju tahap <i>action</i>. • Relevansi: Landasan teoritis untuk menganalisis bagaimana kemudahan akses fitur interaktif (seperti <i>business tag</i> dan CTA pada caption) berfungsi sebagai <i>user-friendly tool</i> yang memicu konversi tindakan audiens dalam klaster penelitian.

Tabel 2.1 Perbandingan dengan Penelitian Terdahulu (Lanjutan)

(1)	(2)	(3)	(4)
11.	Mirantika & Rijanto (2023) [15]	Comparative Analysis of K-Means and K-Medoids Algorithms in Determining Customer Segmentation Using RFM Model	- Metode: CRISP-DM, <i>K-Means</i>, <i>K-Medoids</i>, <i>Elbow Method</i>, dan evaluasi <i>Davies-Bouldin Index</i> (DBI) pada model <i>Recency-Frequency-Monetary</i> (RFM). - Variabel: <i>Recency</i> (kesegaran transaksi), <i>Frequency</i> (frekuensi transaksi), dan <i>Monetary</i> (nilai transaksi). - Hasil: Algoritma <i>K-Means</i> menunjukkan performa lebih unggul dibandingkan <i>K-Medoids</i> berdasarkan nilai DBI yang lebih rendah (lebih optimal) pada ketiga dataset penjualan. - Relevansi: Landasan metodologi komparasi algoritma <i>clustering</i> serta penggunaan metode evaluasi internal untuk menentukan algoritma terbaik dalam pengelompokan pola pemasaran/konsumen.
12.	Yosia & Siregar Bakti (2024) [16]	COMPARATIVE ANALYSIS OF K-MEANS AND K-MEDOIDS ALGORITHMS FOR PRODUCT SALES CLUSTERING AND CUSTOMER	- Metode: Komparasi <i>K-Means</i> dan <i>K-Medoids</i> dengan reduksi dimensi <i>Principal Component Analysis</i> (PCA) pada data transaksi historis. - Variabel: Metrik penjualan produk, data transaksi pelanggan, dan distribusi ruang <i>cluster</i>. - Hasil: Algoritma <i>K-Medoids</i> terbukti lebih unggul dibanding <i>K-Means</i> dalam hal pemisahan klaster (<i>cluster separation</i>) dan kemudahan interpretasi (<i>interpretability</i>) pada data penjualan bisnis. - Relevansi: Rujukan utama mengenai efektivitas algoritma <i>K-Medoids</i> pada pemetaan data pemasaran/bisnis serta acuan penggunaan komparasi untuk optimasi strategi pemasaran berbasis data.

2.2 Rangkuman

Penelitian Lubis dan Iqbal [5] menjadi acuan dalam menerapkan tahapan kerangka kerja *Knowledge Discovery in Databases* (KDD) untuk optimasi media sosial. Studi tersebut menganalisis variabel hari unggah (kode 0-6), jam unggah (kode 0-23), dan metrik interaksi terhadap 29.999 data sampel, dengan temuan kecenderungan waktu unggah optimal pada hari Sabtu pukul 09:00 WIB. Meskipun pengujian dalam literatur tersebut menggunakan data sintesis, alur metodologi *K-Means Clustering* dan *Min-Max Normalization* yang diterapkan memberikan gambaran metodologis yang terstruktur. Bagi penelitian ini, model tersebut diadopsi sebagai dasar pertimbangan dalam menguji rumus jarak Euclidean dan penentuan titik *centroid* guna membantu memetakan kecenderungan karakteristik elemen caption serta waktu distribusi konten pada akun objek penelitian.

Penelitian Nurhidayat dkk. [6] menjadi acuan dalam memetakan karakteristik pengguna media sosial melalui algoritma *K-Means Clustering*, *Elbow Method* dan *Silhouette Coefficient*. Studi terhadap 553 responden ini menganalisis variabel perilaku, preferensi platform, frekuensi akses, dan segmentasi komunikasi untuk memetakan kelompok audiens aktif dan pasif. Bagi penelitian ini, alur tersebut diadopsi sebagai dasar penerapan *Silhouette Coefficient* untuk memvalidasi akurasi jumlah klaster (K) optimal. Selain itu, temuan mengenai dominasi video pendek dalam menarik perhatian menjadi landasan untuk memasukkan variabel Tipe Konten dengan pengodean kategori (1-3) sebagai karakteristik kontekstual guna menghindari bisa kalkulasi jarak pada rumus *K-Means*.

81 Penelitian Ambarsari dkk. [7] digunakan sebagai referensi dalam menganalisis aktivitas digital menggunakan metode *K-Means Clustering* dan *Elbow Method* pada 327 data log riil. Studi tersebut mengamati variabel impresi, jumlah klik, kueri pencarian, dan tingkat ketertarikan, dengan hasil mengindikasikan adanya empat karakteristik audiens, termasuk kelompok minat pasif yang bercirikan impresi tinggi namun klik rendah. Bagi penelitian ini, alur tersebut diadopsi sebagai acuan dalam mempertimbangkan penggunaan data log riil media sosial guna membantu mengevaluasi respons audiens berdasarkan metrik interaksi publik, seperti *likes* dan komentar, pada akun objek penelitian.

Penelitian Kedi dkk. [8] digunakan sebagai referensi pendukung untuk melihat pentingnya adaptasi teknologi dari perspektif manajemen bisnis. Melalui pendekatan *Review Paper*, studi tersebut mengamati variabel adopsi digital, kampanye pemasaran media sosial, dan keberlanjutan bisnis UMKM, dengan kesimpulan mengindikasikan bahwa adaptasi teknologi digital esensial bagi keberlanjutan usaha di tengah keterbatasan teknis. Bagi penelitian ini, literatur tersebut diadopsi sebagai landasan teoritis mengenai urgensi pemanfaatan *Machine Learning* pada skala bisnis kecil, sekaligus menjadi dasar motivasi dalam mempertimbangkan implementasi algoritma *K-Means* pada akun objek penelitian.

Penelitian Nallasivam dkk. [9] digunakan sebagai referensi dalam menganalisis perilaku konsumen melalui pendekatan *mixed-methods* yang mengombinasikan statistik inferensial, *sentiment analysis*, serta algoritma *Machine Learning*. Studi tersebut mengamati metrik keterlibatan Instagram, jenis konten, faktor psikologis, demografi, dan adopsi produk, dengan hasil

mengindikasikan bahwa keterlibatan permukaan (*scrolling* atau sekedar melihat iklan) cenderung kurang signifikan dibandingkan peran rekomendasi personal serta kepercayaan merek berbasis konten interaktif. Bagi penelitian ini, literatur tersebut diadopsi sebagai landasan teoritis mengenai pentingnya fitur interaktif di Instagram, sekaligus menjadi dasar pertimbangan dalam menguji kecenderungan pengaruh elemen *business tag* terhadap tahap tindakan (*Action*) berupa akses transaksi langsung oleh audiens.

Penelitian Nugraha dkk. [10] digunakan sebagai referensi dalam mengamati hubungan antara aktivitas publikasi dan respons audiens melalui pendekatan kuantitatif deskriptif serta analisis isi pada 100 instansi. Studi tersebut mengamati variabel kuantitas unggahan, substansi isi pesan, dan efektivitas interaksi, dengan hasil mengindikasikan bahwa kuantitas konten yang tinggi tidak otomatis menjamin efektivitas interaksi, serta adanya kecenderungan dominasi konten figur pemimpin dibanding informasi publik. Bagi penelitian ini, literatur tersebut diadopsi sebagai dasar argumen dalam menyusun rekomendasi standarisasi konten pada klaster berperforma rendah, sekaligus menjadi acuan dalam menerapkan rumus *Engagement Rate* yang disesuaikan dengan ketersediaan data publik di lapangan melalui pembatasan pada metrik *likes* dan komentar.

Penelitian Nafsyah dkk. [11] digunakan sebagai referensi dalam mengkaji kategorisasi pesan melalui pendekatan kualitatif dan analisis isi berbasis teori Empat Pilar Strategi Media Sosial. Studi tersebut mengamati fokus analisis pada jenis konten dan tingkat interaksi audiens, dengan hasil mengindikasikan bahwa meskipun konten kolaborasi dominan secara kuantitas, jenis konten hiburan cenderung lebih efektif dalam menarik perhatian audiens. Bagi penelitian ini,

literatur tersebut diadopsi sebagai dasar pengelompokkan jenis konten antara promosi usaha dan konten komunitas. Selain itu, alur berpikirnya digunakan sebagai acuan teoritis untuk menganalisis bagaimana perbedaan karakteristik konten tersebut berpotensi memengaruhi penempatan fitur interaktif seperti *business tag* dan fluktuasi nilai *engagement* audiens pada objek penelitian.

Penelitian Yadav dkk. [12] digunakan sebagai acuan teoritis dalam mengkaji penerapan model hierarki komunikasi di media sosial melalui metode analisis konten kualitatif. Studi tersebut mengamati elemen AIDA (*Attention, Interest, Desire, Action*), stimulus visual, emosi, dan personalisasi konten, dengan temuan mengindikasikan bahwa elemen visual kuat berperan memicu *Attention*, sedangkan aspek emosional dan personalisasi cenderung mendorong tahapan *Interest* hingga *Action*. Bagi penelitian ini, literatur tersebut diadopsi sebagai landasan untuk memetakan strategi pada tahap *Action* melalui variabel *Call to Action* (CTA) dan akses transaksi. Selain itu, studi ini digunakan sebagai dasar pertimbangan teoritis dalam memanfaatkan *engagement metrics* (metrik *likes* dan komentar) sebagai indikator performa, sekaligus menjadi acuan utama untuk menganalisis bagaimana karakteristik bahasa promosi yang memanfaatkan unsur emosional serta personalisasi pesan diterapkan pada konten akun objek penelitian.

Penelitian Rahardja dkk. [13] digunakan sebagai referensi dalam mengkaji integrasi antara model komunikasi dan aspek visual melalui pendekatan kualitatif deskriptif berbasis analisis visual serta metrik keterlibatan. Studi tersebut mengamati strategi desain grafis promosional, elemen struktur visual AIDA, dan penempatan *Call to Action* (CTA), dengan kesimpulan mengindikasikan bahwa perancangan visual yang terstruktur dan penempatan CTA yang strategis berpotensi

meningkatkan efektivitas promosi. Bagi penelitian ini, literatur tersebut diadopsi sebagai landasan teoritis guna menganalisis bagaimana struktur visual konten promosi serta penempatan elemen *Action* seperti variabel CTA dan *business tag* memengaruhi kecenderungan pembentukan klaster perperforma interaksi audiens pada akun objek penelitian.

Penelitian Sarfraz dkk. [14] digunakan sebagai acuan teoritis dalam mengkaji proses pengambilan keputusan konsumen melalui metode kualitatif dengan pendekatan analisis tematik berbasis model AIDA. Studi tersebut mengamati variabel keputusan, pengetahuan, dan pengalaman konsumen, dengan hasil mengindikasikan bahwa personalisasi konten dapat meningkatkan minat (*interest*), sementara ketersediaan alat keterlibatan yang praktis (*user-friendly engagement tools*) cenderung mendorong audiens menuju tahap tindakan (*action*). Bagi penelitian ini, literatur tersebut diasopsi sebagai landasan teoritis untuk menganalisis bagaimana kemudahan akses fitur interaktif seperti penempatan *business tag* dan instruksi CTA bagi *caption* berfungsi sebagai instrumen praktis yang memicu konversi tindakan audiens dalam klaster performa yang terbentuk.

Penelitian Mirantika dan Rijanto [15] digunakan sebagai referensi utama dalam mengkaji komparasi algoritma *clustering* untuk segmentasi konsumen menggunakan kerangka kerja CRISP-DM dan model *Recency-Frequency-Monetary* (RFM). Studi tersebut mengamati variabel kesegaran transaksi (*recency*), frekuensi transaksi (*frequency*), dan nilai transaksi (*monetary*) pada tiga dataset penjualan, dengan hasil pengujian menunjukkan bahwa algoritma *K-Means* memiliki performa lebih unggul dibandingkan *K-Medoids* berdasarkan nilai evaluasi *Davies-Bouldin Index* (DBI) yang lebih rendah. Bagi penelitian ini,

alur metodologi tersebut diadopsi sebagai dasar pertimbangan dalam menerapkan teknik komparasi antar-algoritma, serta menjadi acuan dalam memanfaatkan metrik evaluasi internal guna menentukan model pengelompokan yang paling optimal dalam memetakan performa *engagement* audiens.

Penelitian Siregar dkk. [16] digunakan sebagai acuan metodologis dalam mengevaluasi efektivitas algoritma *K-Means* dan *K-Medoids* untuk pemetaan data transaksi dan pemasaran bisnis. Studi tersebut mengombinasikan teknik reduksi dimensi *Principal Component Analysis* (PCA) dengan algoritma *K-Medoids* pada data penjualan historis, dengan temuan mengindikasikan bahwa *K-Medoids* menunjukkan kinerja yang lebih unggul dibanding *K-Means* dalam hal pemisahan kluster (*cluster separation*) dan kemudahan interpretasi (*interpretability*). Bagi penelitian ini, literatur tersebut diadopsi sebagai landasan teoritis mengenai keunggulan algoritma berbasis *medoid* yang lebih tahan terhadap data pencilan (*outliers*), sekaligus menjadi rujukan utama dalam menyusun analisis komparasi guna merumuskan rekomendasi strategi komunikasi konten berbasis data pada objek penelitian.

2.3 Kesimpulan dan Kebaruan Penelitian

Kajian literatur terdahulu mengindikasikan bahwa efektivitas pemasaran media sosial ditentukan oleh kombinasi antara format penyajian media, karakteristik pesan, dan ketepatan waktu distribusi konten. Penggunaan metode *clustering* yang divalidasi dengan *Silhouette Coefficient* dinilai mampu memetakan karakteristik performa tersebut secara terstruktur dan objektif, khususnya dalam menguji ketahanan algoritma seperti *K-Means* dan *K-Medoids* terhadap variansi data *engagement*. Dari aspek komunikasi, integrasi model AIDA

menunjukkan bahwa kombinasi stimulasi elemen visual melalui Tipe Konten, gaya bahasa promosi, serta penempatan fitur interaktif seperti *business tag* dan instruksi *Call to Action* (CTA) pada *caption* berpotensi mendorong tahapan tindakan (*Action*) audiens. Secara manajerial, adopsi *data mining* berbasis komparasi ini diharapkan dapat menjadi landasan bagi UMKM dalam menyusun strategi pemasaran berbasis data.

Kebaruan (*novelty*) dalam penelitian ini dirancang untuk menjembatani celah riset (*research gap*) dari studi-studi terdahulu dengan mengintegrasikan analisis konten berbasis elemen AIDA sekuensial secara kualitatif dengan pendekatan komputasional melalui komparasi algoritma *K-Means* dan *K-Medoids*. Sebagai luaran aplikatif, hasil pengelompokkan dari algoritma terbaik hasil evaluasi ditransformasikan ke dalam model komputasi sederhana pada lembar kerja digital. Hasil ini difungsikan sebagai instrumen teknis yang diharapkan dapat membantu pemilik akun untuk mengamati kecenderungan kinerja unggahan secara mandiri.

Tabel 2.6 Perbandingan Research Gap

Dimensi Perbandingan	Penelitian Terdahulu	Penelitian Ini
Fokus Analisis	Umumnya terpisah antara aspek teknis komputasional saja atau teori komunikasi saja.	Terintegrasi antara metode komputasional komparatif (<i>K-Means</i> vs <i>K-Medoids</i>) dan teori komunikasi pemasaran (Model AIDA).
Sumber Data	Dominan menggunakan data sintesis atau studi kasus pada korporasi atau instansi besar.	Menggunakan data riil historis publik akun media sosial UMKM lokal yang rentan terhadap <i>outliers</i> .
Metode Validasi	Mayoritas menitikberatkan pada satu algoritma tunggal (<i>K-Means</i>) atau penentuan klaster melalui <i>Elbow Method</i> .	Menitikberatkan pada studi komparasi performa dua algoritma (<i>K-Means</i> vs <i>K-Medoids</i>) dengan pengujian <i>Silhouette Coefficient</i> untuk menentukan algoritma dengan akurasi dan kepresisian terbaik.
Output	Bersifat generalisasi	Menghasilkan rekomendasi pilar komunikasi,

Dimensi Perbandingan	Penelitian Terdahulu	Penelitian Ini
Strategi	untuk cakupan industri secara luas dan jarang menyertakan instrumen implementasi untuk objek lokal.	kombinasi format tipe konten, manajemen waktu publikasi (<i>Tailored Strategy</i>) berbasis kluster algoritma terbaik, serta prototipe alat evaluasi berbasis lembar kerja digital.

Berdasarkan matriks perbandingan tersebut, aspek kebaruan yang diusulkan dalam penelitian ini meliputi:

1. Integrasi Komparasi Komputasional dan Analisis Konten (Double-Layer Analysis): Penelitian ini mengombinasikan pengujian teknis berupa komparasi algoritma K-Means dan K-Medoids yang divalidasi dengan Silhouette Coefficient, dengan analisis karakteristik isi berbasis model AIDA. Pendekatan ini diharapkan dapat memberikan gambaran performa konten secara menyeluruh serta membuktikan ketahanan algoritma terhadap data pencilan (outliers).
2. Kontekstualisasi Sektor UMKM Lokal: Berbeda dengan mayoritas studi terdahulu yang mengamati skala korporasi besar atau instansi publik, riset ini menempatkan UMKM lokal sebagai objek pengamatan utama. Fokus tersebut bertujuan untuk mengkaji potensi implementasi teknik data mining sebagai instrumen pendukung keputusan yang terjangkau bagi pelaku usaha kecil.
3. Eksplorasi Pola Publikasi, Karakteristik Pesan, dan Elemen Interaktif: Penelitian ini mengamati hubungan antara aktivitas unggahan dan efektivitas interaksi riil audiens berdasarkan metrik likes dan comments. Analisis juga diarahkan untuk melihat bagaimana pemilihan tipe konten (format media), pendekatan bahasa promosi (unsur emosional dan personalisasi), waktu publikasi, serta pemanfaatan fitur interaktif seperti business tag memengaruhi

kecenderungan tingkat keterlibatan (engagement) publik berdasarkan karakteristik klaster algoritma terbaik yang terbentuk.

78

BAB III

METODE PENELITIAN

1

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.

6

1

76

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

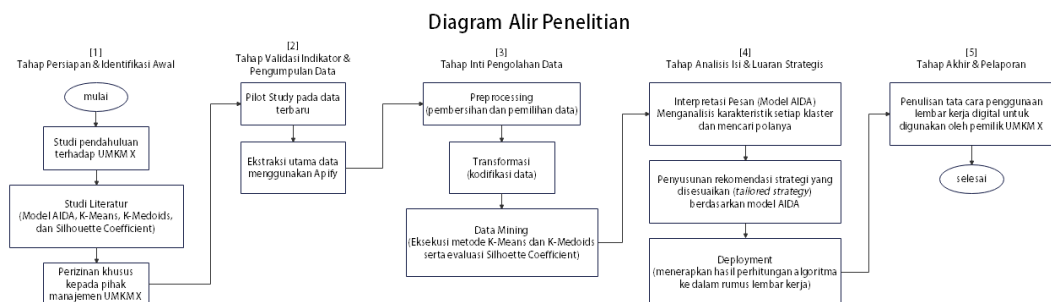
13

1

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

1 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.

1 3.1.1 Subjek, Objek, dan Justifikasi Pemilihan Data

1 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.

1 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	Yang mau
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	Salah si
8	4/3/2026	6	yes	yes	yes	yes	22	1	promotional	carousel	Yang mau
9	26/2/2026	1	yes	yes	yes	yes	17	1	promotional	carousel	Pesan h
10	25/2/2026		yes	yes	yes	yes	94	1	promotional	carousel	Rekom
11	25/2/2026	2	yes	yes	yes	yes	53	1	promotional	carousel	Paket R
12	23/2/2026		yes	yes	yes	yes	17	1	promotional	carousel	TEH TA
13	23/2/2026	1	yes	yes	yes	yes	84	1	promotional	carousel	Paket Bul
14	22/2/2026		yes	yes	yes	yes	39	1	promotional	carousel	PAKET
15	22/2/2026	1	yes	yes	yes	yes	7	1	promotional	carousel	Hamper
16	21/2/2026		yes	yes	yes	yes	89	1	promotional	carousel	Cake tir
17	21/2/2026		yes	yes	yes	yes	45	1	promotional	carousel	Cemilar
18	21/2/2026	1	yes	yes	yes	yes	11	1	promotional	carousel	Yang ma
19	20/2/226	1	yes	yes	yes	yes	43	3	promotional	carousel	Paket Bul
20	19/2/2026		yes	yes	no	no	68	1	possibly	carousel	Paket bi

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	<i>Call To Action</i> (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	<i>Action</i>	Penyebutan (mention/tag)	Variabel	Diadopsi dari

1

	Tag		akun bisnis lain atau mitra kolaborasi pada unggahan. Skala biner: 0 = Tidak Ada, 1 = Ada.	Karakteristik (Kontekstual Klaster)	kemudahan konversi tahap <i>action</i> [13].
--	-----	--	--	-------------------------------------	--

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 klaster 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	Jam 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 = Single Photo / Image 2 = Carousel 3 = Reels / Video	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 klaster 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.

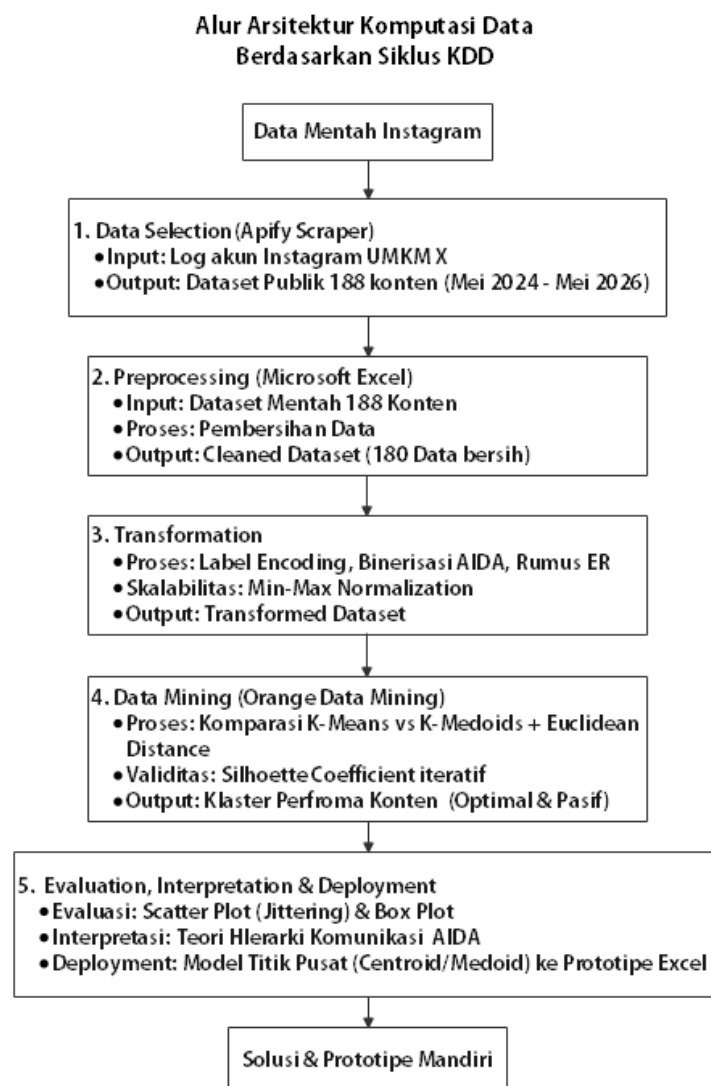
3.2 Perangkat Pendukung Penelitian

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, Ci*) pada algoritma *K-Means*, atau objek rill (*Medoid, Mi*) 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-kluster 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 kluster 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 kluster 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 kluster:

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

Keterangan:

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

Instrumen teknis interaktif ini memungkinkan pemilik akun UMKM X mengevaluasi kinerja konten baru secara mandiri dan berkelanjutan.

52

BAB IV

HASIL DAN PEMBAHASAN

1

Bab ini menyajikan hasil pelaksanaan setiap tahapan penelitian berbasis kerangka kerja *Knowledge Discovery in Databases* (KDD), mencakup hasil pengolahan data, analisis dan evaluasi klaster secara komprehensif, serta rangkuman temuan utama penelitian.

4.1 Hasil Pengolahan Data

Subbab ini memaparkan pengumpulan, pembersihan, dan pemetaan data awal. *Web scraping* via Apify menghasilkan 188 unggahan publik akun UMKM X (Mei 2024–Mei 2026). Tahap *data cleaning* mengeliminasi 8 data *noise*, menghasilkan 180 data bersih. *Engagement Rate* dihitung menggunakan pembagi konstan 15.477 *followers*.

Aktivitas publikasi pada tahun 2024 cenderung stabil, meski mengalami penurunan pada bulan Oktober dan Desember. Distribusi frekuensi unggahan tahun 2024 disajikan pada Tabel 4.1.

65

Tabel 4.1 Distribusi Fekuensi Unggahan UMKM X Tahun 2024

Bulan	Jumlah Post/Bulan	Rata-Rata Postingan / Bulan
Mei	12	1.5
Juni	13	1.625
Juli	9	1.125
Agustus	9	1.125
September	11	1.375
Oktober	6	0.75
November	8	1
Desember	5	0.625

Memasuki tahun 2025, frekuensi unggahan mengalami penurunan dibanding tahun sebelumnya. Rata-rata unggahan bulanan yang berada di atas

angka 1 hanya tercatat pada bulan September. Rincian distribusi unggahan sepanjang tahun 2025 dirangkum pada Tabel 4.2.

Tabel 4.2 Distribusi Frekuensi Unggahan UMKM X Tahun 2025

Bulan	Jumlah Post/Bulan	Rata-Rata Postingan / Bulan
Januari	3	0.25
Februari	7	0.58333333
Maret	8	0.66666667
April	1	0.08333333
Mei	11	0.91666667
Juni	3	0.25
Juli	8	0.66666667
Agustus	7	0.58333333
September	12	1
Oktober	8	0.66666667
November	4	0.33333333
Desember	4	0.33333333

Sementara itu, pada sisa periode pengamatan di awal tahun 2026, aktivitas publikasi menunjukkan kembali tren peningkatan, khususnya pada bulan Februari. Data frekuensi unggahan tahun 2026 ditunjukkan pada Tabel 4.3.

Tabel 4.3 Distribusi Frekuensi Unggahan UMKM X Tahun 2026

Bulan	Jumlah Post/Bulan	Rata-Rata Postingan / Bulan
Januari	1	0.25
Februari	19	4.75
Maret	7	1.75
April	4	1

Fluktuasi intensitas unggahan ini menjadi basis evaluasi hubungan antara frekuensi publikasi dan *Engagement Rate* pada tahap berikutnya.

Setelah sebaran frekuensi dipetakan, potongan sampel dari 180 data bersih disajikan pada Gambar 4.1.

caption	likesComments	commerce	businessTag	promoLanguage	CTA_Presence	transactionAccess	date	type	postTime
Nastar J	27	0	iya	iya	iya	iya	26 April 2026	Carousel	17:15:23.000Z
Hi guys!	19	0	iya	iya	iya	iya	18 April 2026	Carousel	01:45:14.000Z
Hi guys!	91	0	iya	iya	iya	iya	12 April 2026	Carousel	14:01:39.000Z
Hi guys!	75	0	iya	iya	iya	iya	02 April 2026	Carousel	02:37:58.000Z
Bukber	51	0	tidak	iya	iya	tidak	11 Maret 2026	Video	04:40:59.000Z
Paket B	62	0	iya	iya	iya	iya	10 Maret 2026	Carousel	10:17:11.000Z
Salah s	38	2	tidak	iya	iya	iya	07 Maret 2026	Carousel	13:49:51.000Z
Salah s	113	1	iya	iya	iya	iya	06 Maret 2026	Carousel	04:17:30.000Z
Yang m	15	1	iya	iya	iya	iya	06 Maret 2026	Carousel	22:53:36.000Z
Salah s	34	1	iya	iya	iya	iya	04 Maret 2026	Carousel	12:27:17.000Z
Yang m	22	1	iya	iya	iya	iya	03 Maret 2026	Carousel	19:45:32.000Z
Pesan h	17	1	iya	iya	iya	iya	26 Februari 2026	Carousel	16:37:04.000Z
Rekom	96	1	iya	iya	iya	iya	25 Februari 2026	Carousel	15:55:13.000Z
Paket R	53	1	iya	iya	iya	iya	25 Februari 2026	Carousel	13:52:20.000Z

Gambar 4.1 Potongan Lembar Kerja Dataset 2 Tahun (Mei 2024- Mei 2026)

Gambar 4.1 menampilkan 15 sampel data mentah hasil pembersihan. Pada tahap ini, variabel kualitatif (CTA, Business Tag, Bahasa Promosi, Akses Transaksi) masih berformat tekstual ("Iya"/"Tidak"), sedangkan metrik interaksi (likes, comments, tipe konten, waktu unggah) berupa angka murni. Dataset ini kemudian ditransformasikan ke format encoding numerik agar siap diproses oleh algoritma K-Means dan K-Medoids.

4.2 Hasil Transformasi

Tahap ini berfokus pada konversi seluruh variabel ke dalam matriks numerik agar siap diproses oleh algoritma *K-Means* dan *K-Medoids*. Proses transformasi mencakup tiga langkah utama:

1. Pengodean Kategori (*Label Encoding*), mengonversi variabel waktu (hari ke indeks 0–6 dan jam ke 0–23) serta tipe konten (1=Photo, 2=Carousel, 3=Reels).
2. Kodifikasi Biner (*Dummy Variable*), mengubah karakteristik pesan kualitatif (*promoLanguage*, *CTA_Presence*, *transactionAccess*, dan *businessTag*) menjadi nilai biner (1 = ada, 0 = tidak ada).
3. Normalisasi *Engagement Rate* (ER), menghitung nilai ER riil dari metrik *likes* dan *comments*, lalu menyelaraskan skalanya ke rentang desimal [0, 1]

menggunakan *Min-Max Normalization* guna menghindari dominasi bobot metrik saat klusterisasi.

Contoh kalkulasi nilai ER dan proses normalisasinya pada data ke-1 dirumuskan sebagai berikut:

$$Engagement\ Rate = \frac{27 + 0}{15.477} \times 100\% = 0\%$$

Kemudian dilakukan normalisasi dengan rumus *Min-Max*:

$$X_{norm} = \frac{0\% - 0\%}{18\% - 0\%} = 0,00$$

Sistem kodifikasi dan pemetaan peran setiap variabel ke dalam format angka dirangkum pada Tabel 4.4.

Tabel 4.4 Aturan Pengodean (*Encoding*) Transformasi Variabel

Variabel	Kategori Asli (Tekstual/Mentah)	Kode Numerik Hasil Desain	Role di Orange
Hari_unggah	Nama hari tekstual (Senin-Minggu)	0 s.d 6	Meta
Jam_post	Waktu rilis format jam digital (00:00-23:59)	0 s.d 23	Meta
Type	Jenis format visual	1 = Photo 2 = Carousel 3 = Reels	Meta
promoLanguage	Kalimat penawaran / diskon	0 = tidak ada 1 = ada	Meta
CTA_Presence	Tombol / interaksi ajakan	0 = tidak ada 1 = ada	Meta
TransactionAccess	Link, kontak / cara pembelian produk	0 = tidak ada 1 = ada	Meta
BusinessTag	Tag akun bisnis mitra	0 = tidak ada 1 = ada	Meta
ER_normalisasi	Rasio interaksi (<i>Engagement</i>)	Desimal fraksional (0.00 s.d 1.00)	Feature

Penerapan aturan pengodean tersebut menghasilkan matriks numerik utuh yang siap dianalisis. Potongan sampel dataset hasil transformasi disajikan pada Gambar 4.2.

apf	ssTagged	Language	Presence	tionAcces	Hari_ungg	type	Jam_post	Engagement Rate	ER_normalisasi	
Nast	1	1	1	1	6	2	17	0%	0.00	htt
Hi gu	1	1	1	1	5	2	1	0%	0.00	htt
Hi gu	1	1	1	1	6	2	14	1%	0.00	htt
Hi gu	1	1	1	1	3	2	2	0%	0.00	htt
Bukb	0	1	1	0	2	3	4	0%	0.00	htt
Pake	1	1	1	1	1	2	10	0%	0.00	htt
	0	1	1	1	5	2	13	0%	0.00	htt
Salah	1	1	1	1	4	2	4	1%	0.00	ht
Yang	1	1	1	1	4	2	22	0%	0.00	htt
Salah	1	1	1	1	2	2	12	0%	0.00	htt
Yang	1	1	1	1	1	2	19	0%	0.00	htt
Pesai	1	1	1	1	3	2	16	0%	0.00	http
Reko	1	1	1	1	2	2	15	1%	0.00	http
Pake	1	1	1	1	2	2	13	0%	0.00	htt

Gambar 4.2 Potongan Dataset Hasil Transformasi Matriks Numerik

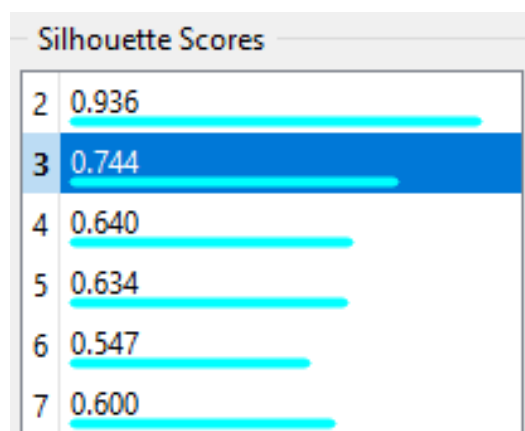
Sebagaimana terlihat pada Gambar 4.2, seluruh variabel kualitatif dan metrik interaksi telah berformat matriks numerik. Variabel berlabel *Meta* berfungsi sebagai atribut konteks pascaklasterisasi, sedangkan variabel *ER_normalisasi* berlabel *Feature* menjadi data masukan utama yang diproses oleh algoritma *K-Means* dan *K-Medoids* pada tahap selanjutnya.

4.3 Hasil Implementasi

Di bawah ini adalah penjelasan dari hasil penerapan dari tahapan di atas.

1. Penentuan Jumlah Kluster (K) Optimal dan Evaluasi Komparatif

Penentuan jumlah kluster optimal dilakukan dengan mengevaluasi nilai *Silhouette Coefficient* pada rentang K=2 hingga K=5. Nilai *Silhouette Scores* untuk algoritma K-Means dari perangkat lunak Orange Data Mining ditunjukkan pada Gambar 4.3.



Gambar 4.3 Nilai Silhouette Scores K-Means untuk Berbagai Nilai K

15

Berdasarkan Gambar 4.3, pembentukan $K=2$ menghasilkan nilai *Silhouette Coefficient* tertinggi sebesar 0,936. Namun, secara substansi analisis bisnis, pembagian dua kelompok cenderung terlalu umum (*dikotomis*) karena hanya memisahkan data secara ekstrem menjadi kelompok performa tinggi dan rendah tanpa mengakomodasi kelompok performa sedang. Oleh karena itu, ditetapkan parameter $K=3$ dengan skor *Silhouette* sebesar 0,744. Pemilihan ini didasarkan pada kebutuhan interpretasi bisnis (*business interpretability*) agar variasi performa konten UMKM X terpetakan secara komprehensif ke dalam tiga segmen (rendah, sedang, dan tinggi), dengan kualitas pengelompokan yang tetap berada pada kategori struktur kuat (*strong structure*).

74

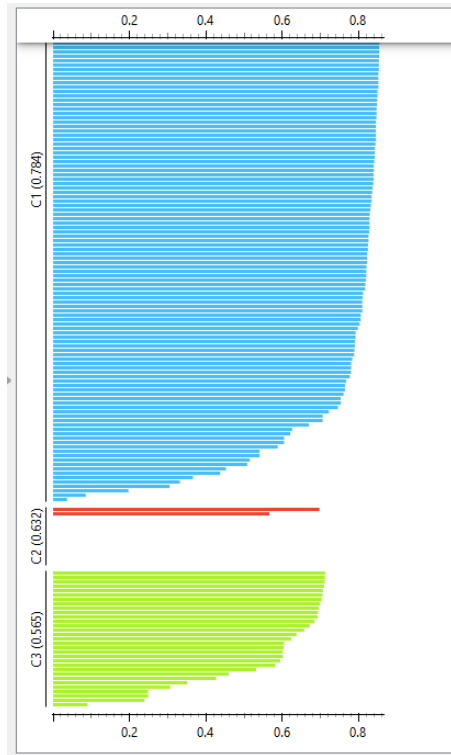
1

Pada algoritma K-Medoids, penentuan dan evaluasi pembentukan kluster dilakukan melalui pengamatan *Silhouette Plot* secara visual serta evaluasi skor rata-rata *Silhouette* total. Berdasarkan hasil pengujian $K=3$, algoritma K-Medoids menghasilkan skor *Silhouette Coefficient* sebesar 0,621, yang mengindikasikan bahwa struktur pengelompokannya berada pada kategori struktur yang baik (*reasonable structure*).

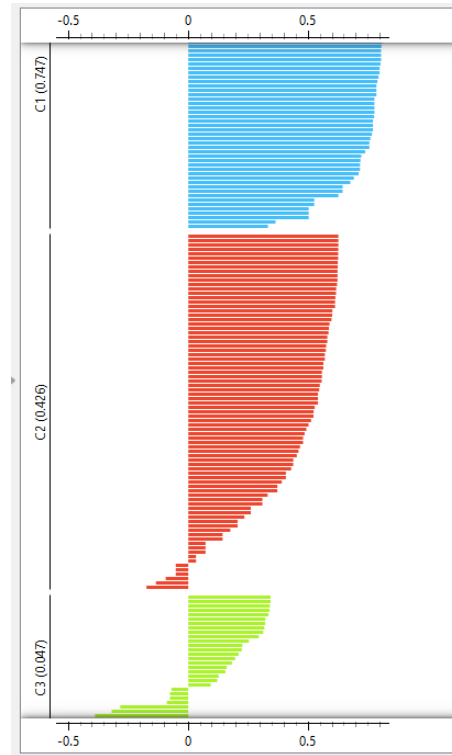
16

11

Guna memperdalam perbandingan kinerja antara K-Means dan K-Medoids pada $K=3$, evaluasi dilakukan secara visual melalui sebaran *Silhouette Plot* untuk masing-masing algoritma pada Gambar 4.4.



Gambar 4.4a Silhouette Plot K-Means



Gambar 4.4a Silhouette Plot K-Means

Gambar 4.4 Visualisasi *Silhouette Plot* (a) K-Means dan (b) K-Medoids (K=3)

Berdasarkan Gambar 4.4, terlihat perbedaan karakteristik pembentukan kluster yang signifikan antara kedua algoritma. Pada K-Means (Gambar 4.4a), seluruh nilai Silhouette tiap kelompok bernilai positif, yaitu C1 sebesar 0,784, C2 sebesar 0,632, dan C3 sebesar 0,565. Hal ini menandakan bahwa setiap data teralokasi secara konsisten tanpa ada indikasi kesalahan pengelompokan. Sementara itu, pada K-Medoids (Gambar 4.4b), meskipun klaster C1 menunjukkan kerapatan yang tinggi (0,747), terdapat sebagian kecil data pada C2 (0,426) dan C3 (0,047) yang bernilai negatif. Adanya nilai negatif tersebut mengindikasikan keberadaan data yang berada di area transisi (*overlapping*) antar-klaster.

Dengan demikian, hasil visualisasi ini makin memperkuat alasan pemilihan K-Means sebagai model utama dalam analisis segmentasi konten, sedangkan K-Medoids digunakan sebagai alat validasi pembandingan.

2. Sampel Distribusi Data Hasil Klasterisasi K-Means

Algoritma K-Means mengelompokkan 180 data unggahan berdasarkan kesamaan jarak fitur (*feature similarity*). Sampel distribusi data hasil klasterisasi ditunjukkan pada Tabel 4.5.

Tabel 4.5 Sampel Pengelompokan Data Menggunakan Algoritma K-Means

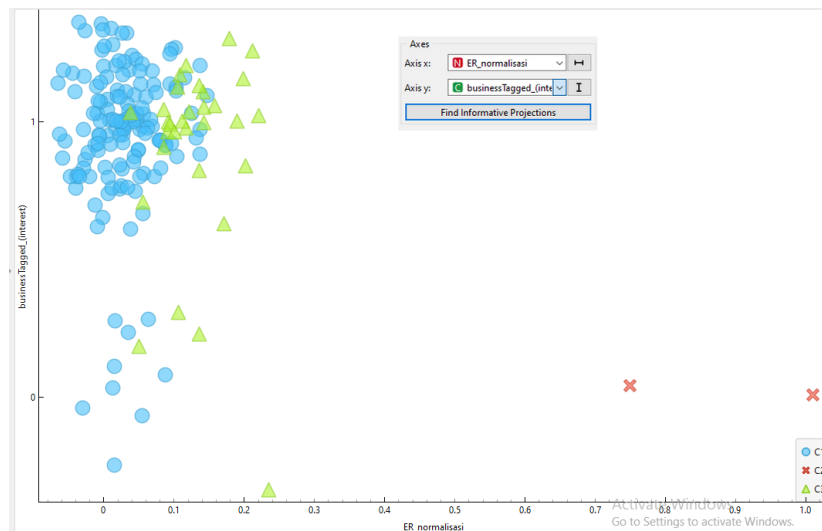
No	Tipe konten	Hari unggah	Jam posting	CTA	Bahasa promosi	Akses transaksi	Tag bisnis	Engagement Rate (ER)	klaster
1	2	6	15	1	1	1	1	0.885184	C1
2	3	2	17	1	1	1	0	0.626736	C1
3	3	1	14	0	0	0	0	17.8071	C2
4	3	2	2	0	0	0	0	13.1873	C2
5	2	2	16	1	1	1	0	4.16747	C3
6	3	2	14	1	1	0	1	2.55217	C3

Berdasarkan sampel pada Tabel 4.5, kecenderungan umum dari setiap kelompok teridentifikasi sebagai berikut:

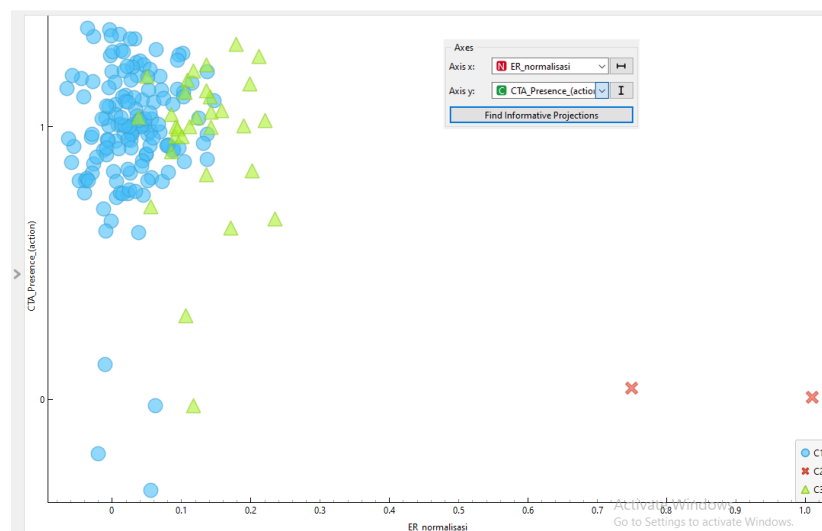
- 1) Klaster 1 (C1): Cenderung menggunakan seluruh fitur pesan secara eksplisit (*CTA*, *promo*, *akses transaksi*, dan *tag bisnis* bernilai 1), namun menghasilkan *Engagement Rate* (ER) pada rentang rendah (di bawah 1%).
- 2) Klaster 2 (C2): Mayoritas tidak menerapkan elemen pesan formal (fitur teks bernilai 0), tetapi mencatatkan nilai ER tertinggi di dalam dataset (misal: 17,80% dan 13,18%).
- 3) Klaster 3 (C3): Menggunakan elemen pesan secara parsial/selektif, dengan sebaran nilai ER berada pada skala moderat (2%–4%).

3. Analisis Sebaran Fitur Konten (*Scatter Plot*)

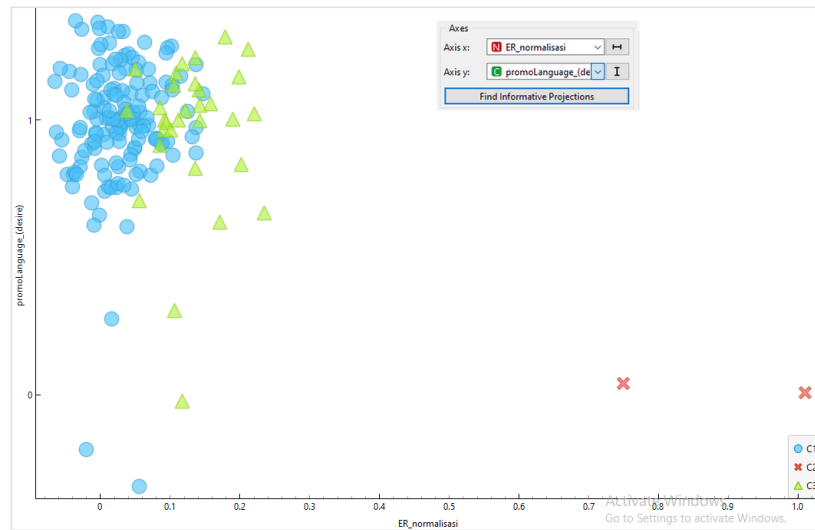
Pola hubungan antara *Engagement Rate* ternormalisasi (sumbu X) dan masing-masing fitur kualitatif karakteristik pesan dipetakan melalui diagram *Scatter Plot* pada Gambar 4.5 hingga Gambar 4.10.



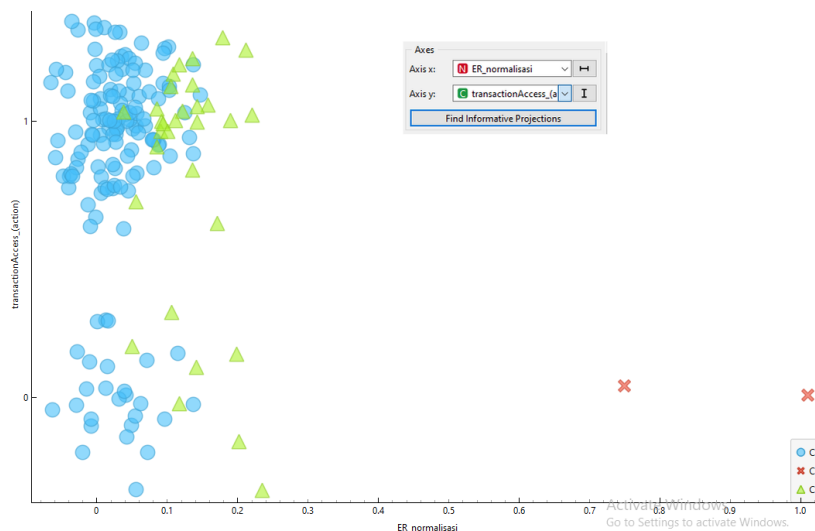
Gambar 4.5 Visualisasi Kluster Melalui Grafik Scatter Plot (bisnisTag)



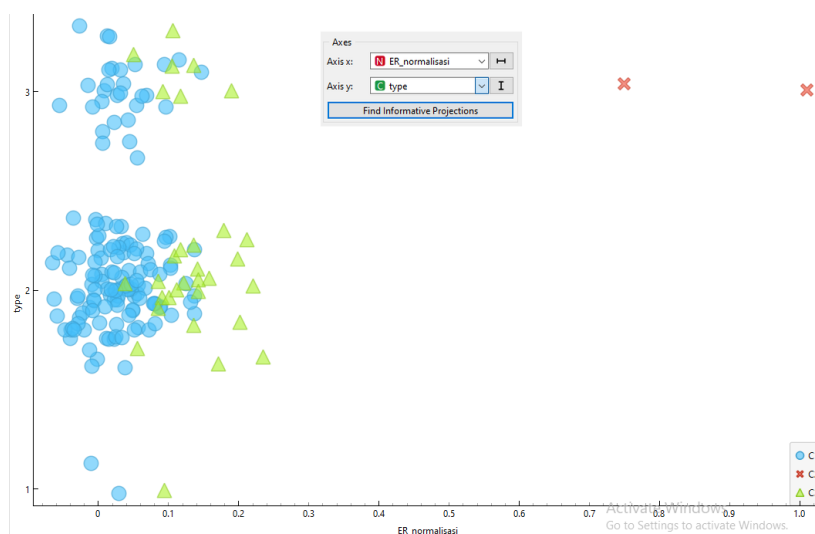
Gambar 4. 6 Visualisasi Kluster Melalui Grafik Scatter Plot (CTA)



Gambar 4. 7 Visualisasi Kluster Melalui Grafik Scatter Plot (PromoLang)



Gambar 4. 8 Visualisasi Kluster Melalui Grafik Scatter Plot (AksesTransaksi)



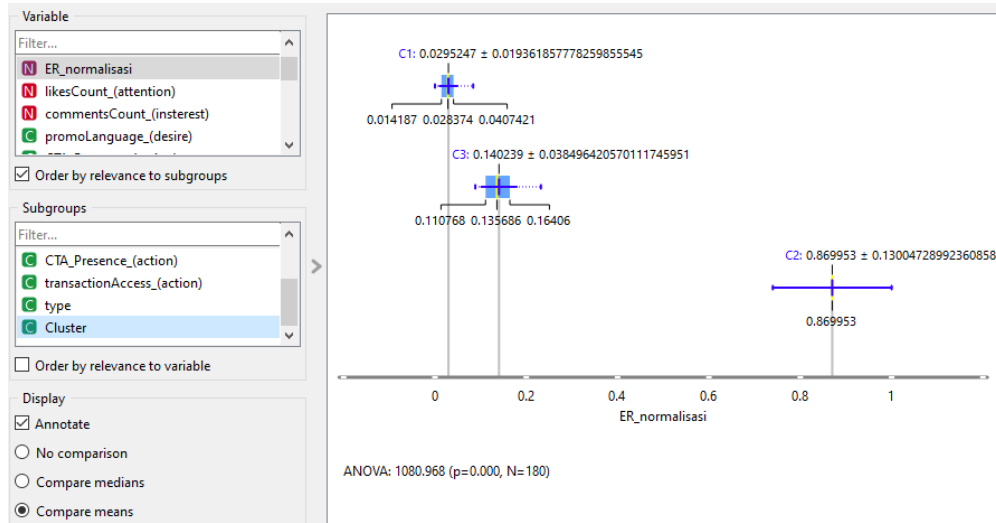
Gambar 4.9 Visualisasi Kluster Melalui Grafik Scatter Plot (TipeKonten)

Berdasarkan Gambar 4.5 hingga 4.10, grafik *Scatter Plot* menunjukkan pola sebaran yang konsisten dan terpisah secara tegas antar-klaster pada seluruh variabel yang diuji:

- 1) Klaster 1 / C1 (Lingkaran Biru): Memusat padat di sisi kiri dengan Engagement Rate (ER) mendekati 0,00. Hal ini mengindikasikan bahwa penerapan fitur pesan transaksional secara eksplisit (businessTag, CTA, promoLanguage, transactionAccess) belum berdampak signifikan terhadap peningkatan interaksi audiens.
- 2) Klaster 2 / C2 (Silang Merah): Terisolasi di pojok kanan dengan ER tertinggi (0,70 - 1,00). Meskipun tidak menggunakan instruksi teks formal (fitur biner bernilai 0), seluruh anggota C2 berada pada format video (Gambar 4.10), membuktikan bahwa lonjakan interaksi dipicu dominan oleh daya tarik visual video pendek (Reels).
- 3) Klaster 3 / C3 (Segitiga Hijau): Tersebar di area tengah dengan skala ER moderat (0,10 - 0,20). Kombinasi penggunaan pesan secara selektif (soft-selling) dengan variasi format Carousel dan Video terbukti efektif menjaga stabilitas interaksi pada tingkat menengah.

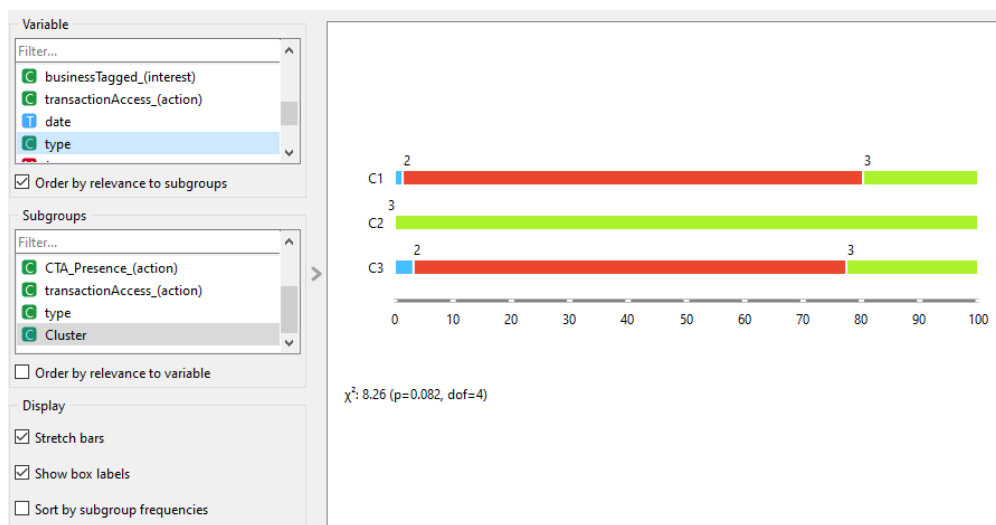
4. Evaluasi Distribusi dan Validasi Statistik Klaster (*Box Plot & Bar Chart*)

Guna memvalidasi perbedaan performa interaksi dan pemakaian format konten antar-kelompok secara kuantitatif, dilakukan analisis sebaran nilai menggunakan *Box Plot* dan proporsi tipe konten melalui *Bar Chart* pada Gambar 4.11 dan Gambar 4.12.



Gambar 4.10 Distribusi *ER* Ternormalisasi per Kluster (Box Plot)

Berdasarkan Gambar 4.11, sebaran rata-rata *Engagement Rate* ternormalisasi menunjukkan pemisahan kelompok yang sangat jelas. Kluster C1 berada pada rentang paling bawah dengan nilai rata-rata 0,0295, Kluster C3 berada pada tingkat moderat dengan nilai rata-rata 0,1402, dan Kluster C2 mencatatkan sebaran tertinggi dengan rata-rata 0,8699. Hasil pengujian hipotesis menggunakan ANOVA menghasilkan nilai $F = 1080,968$ ($p = 0,000 < 0,05$), yang membuktikan bahwa perbedaan tingkat interaksi antar-kluster signifikan secara statistik.



Gambar 4.11 Distribusi Proporsi Tipe Konten pada Masing-Masing Kluster

Sementara itu, Gambar 4.12 memperlihatkan distribusi proporsi tipe konten di setiap kelompok. Klaster C1 dan C3 didominasi oleh format *Carousel* (kode 2) dengan akumulasi proporsi mencapai 75% - 80%. Sebaliknya, Klaster C2 diisi 100% secara eksklusif oleh format Video (kode 3). Hasil uji *Chi-Square* ($X^2 = 8,26$; $p = 0,082$) secara deskriptif makin menegaskan adanya pemusatan format video pendek pada kelompok berkinerja unggul.

5. Profiling Klaster dan Interpretasi Bisnis

Sebagai tahap akhir dari Subbab 4.3, seluruh temuan statistik dan visualisasi dirangkum ke dalam tabel *profiling* untuk menerjemahkan hasil analisis data menjadi acuan strategi pemasaran bagi UMKM X.

Klaster	Segmen Performa	Karakteristik Performa Konten	Rata-rata ER	Rekomendasi Strategi Bisnis
C1	Low Performance	Penggunaan fitur teks jualan eksplisit tinggi (CTA, Promo, Tag Bisnis). Didominasi format <i>Carousel</i> & Gambar.	0,88%	Reduce/Re-evaluate: Kurangi penyampaian teks <i>hard selling</i> yang kaku dan ubah pendekatan visual menjadi lebih interaktif.
C3	Moderate Performance	Penggunaan elemen pesan secara parsial. Kombinasi seimbang antara <i>Carousel</i> dan Video.	3,35%	Maintain: Dijadikan pilar konten harian (<i>daily feeds</i>) untuk menjaga kesadaran merek (<i>brand awareness</i>) secara stabil.
C2	High Performance	Minim instruksi teks formal. dikuasai 100% oleh format Video Pendek (<i>Reels</i>).	15,50%	Scale Up: Fokus utama untuk menciptakan konten viral (<i>viral driver</i>) dengan memperbanyak video edukatif/bercerita.

4.4 Hasil Evaluasi dan Kontekstualisasi Temuan Model AIDA

Tahap ini menginterpretasikan temuan kuantitatif algoritma *K-Means* ke dalam perspektif kualitatif kerangka kerja pemasaran AIDA (*Attention, Interest, Desire, Action*):

1) Klaster 1 (C1) – Pendekatan *Action* vs. *Attention*:

Didominasi format *Carousel* berorientasi transaksional langsung (*Action*). Namun, rata-rata *Engagement Rate* (ER) yang rendah (0,0295) menunjukkan bahwa pendekatan *hard-selling* dominan kurang optimal dalam membangun daya tarik awal (*Attention*).

2) Klaster 2 (C2) – Mendorong *Organic Interest*:

Didominasi 100% format Video dengan rata-rata ER tertinggi (0,8699). Hal ini membuktikan bahwa daya tarik visual dinamis sangat efektif memicu respons *Attention* dan *Interest*, melebihi efektivitas instruksi transaksional pada teks.

3) Klaster 3 (C3) – Keseimbangan Selektif (*Selective AIDA Balance*):

Didominasi format *Carousel* dengan rata-rata ER moderat (0,1402). Penerapan elemen pesan secara selektif (*soft-selling*) terbukti potensial memelihara minat (*Desire*) dan menjaga stabilitas interaksi audiens.

Temuan ini memberikan arah rekomendasi *data-driven marketing* bagi UMKM X. Pengelolaan konten disarankan beralih dari pendekatan transaksional penuh yang kaku (Klaster 1) dengan memanfaatkan visual dinamis (Klaster 2) untuk membangun *Attention* dan *Interest* di fase awal. Selanjutnya, promosi selektif (Klaster 3) disisipkan secara bertahap untuk mengarahkan audiens dari

fase *Desire* menuju *Action*. Rincian panduan praktis dipaparkan lebih lanjut pada subbab implikasi manajerial.

4.5 Implikasi Operasional dan Pemeliharaan Model

Tahap pemeliharaan menghasilkan prototipe evaluasi performa konten berbasis Microsoft Excel untuk UMKM X yang efisien, mudah dioperasikan, dan sesuai dengan keterbatasan teknis pengelola.

1. Prosedur Kerja Prototipe

Sistem operasionalisasi lembar kerja ini dirancang melalui dua tahapan utama:

- 1) Pengumpulan dan Input Data, Data historis unggahan ditarik berkala (*scraping* minimal sebulan sekali). Admin memasukkan metrik interaksi (*likes*, *comments*) serta karakteristik kualitatif pesan ke kolom input.
- 2) Integrasi Formula Otomatis, Data diproses otomatis melalui dua tahap komputasi:
 - a. *Normalisasi Min-Max*: Menyelaraskan nilai Engagement Rate (ER) baru ke skala [0, 1] berdasarkan parameter model awal.
 - b. Jarak Euclidean: Menghitung jarak geometris nilai ER baru yang telah dinormalisasi (X_{baru}) terhadap titik pusat (*centroid*) setiap kluster (C_i).

$$d = \sqrt{(X_{baru} - C_i)^2}$$

Keterangan

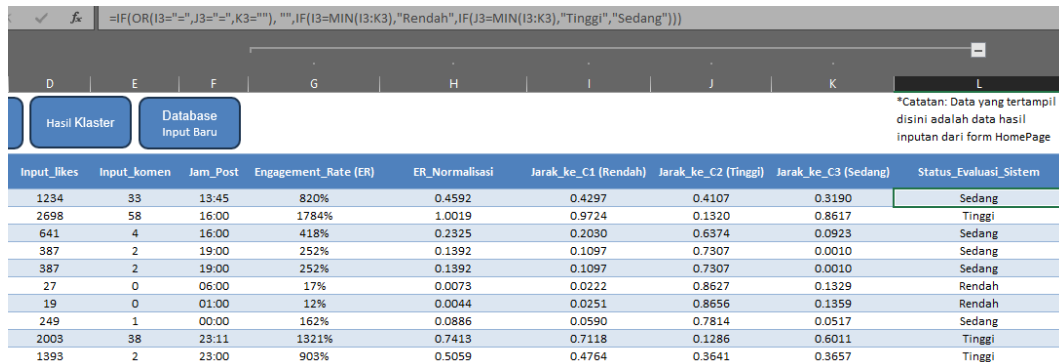
- X_{baru} = Nilai normalisasi data baru.
- C_i = Nilai *centroid* (rata-rata) untuk masing-masing kluster i (C_1 , C_2 , C_3)

- 3) Logika Klasifikasi (Nested IF), Membandingkan jarak Euclidean untuk menentukan jarak terdekat sebagai label performa data baru ("Rendah", "Sedang", atau "Tinggi"). Formula ini dilengkapi logika OR agar sel tetap bersih jika data belum diisi:

=IF(OR(I3="", J3="", K3=""), "", IF(I3=MIN(I3:K3), "Rendah", IF(J3=MIN(I3:K3), "Tinggi", "Sedang")))

Keterangan:

- I3 = Jarak Euclidean ke pusat Kluster 1 (kategori performa “Rendah”)
- J3 = Jarak Euclidean ke pusat Kluster 2 (kategori performa “Tinggi”)
- K3 = Jarak Euclidean ke pusat Kluster 3 (kategori performa “Sedang”)
- MIN(I3:K3) = mencari nilai jarak terkecil.



Input_likes	Input_komen	Jam_Post	Engagement_Rate (ER)	ER_Normalisasi	Jarak_ke_C1 (Rendah)	Jarak_ke_C2 (Tinggi)	Jarak_ke_C3 (Sedang)	Status_Evaluasi_Sistem
1234	33	13:45	820%	0.4592	0.4297	0.4107	0.3190	Sedang
2698	58	16:00	1784%	1.0019	0.9724	0.1320	0.8617	Tinggi
641	4	16:00	418%	0.2325	0.2030	0.6374	0.0923	Sedang
387	2	19:00	252%	0.1392	0.1097	0.7307	0.0010	Sedang
387	2	19:00	252%	0.1392	0.1097	0.7307	0.0010	Sedang
27	0	06:00	17%	0.0073	0.0222	0.8627	0.1329	Rendah
19	0	01:00	12%	0.0044	0.0251	0.8656	0.1359	Rendah
249	1	00:00	162%	0.0886	0.0590	0.7814	0.0517	Sedang
2003	38	23:11	1321%	0.7413	0.7118	0.1286	0.6011	Tinggi
1393	2	23:00	903%	0.5059	0.4764	0.3641	0.3657	Tinggi

Gambar 4. 12 Tampilan Prototipe Lembar Kerja Pemeliharaan Konten dan Simulasi Otomatisasi Kluster

Berikut adalah cara pengujian dan validasi prototipe dilakukan pada Gambar 4.11.

- 1) Skenario A (Data Historis): Pengujian data aktual konsisten menghasilkan label kluster yang sesuai dengan model awal, mengonfirmasi keakuratan formula.
- 2) Skenario B (Simulasi Data Acak/Batas Kritis): Pengujian data acak membuktikan formula tetap stabil dan sensitif dalam mengatribusikan label berdasarkan jarak terdekat.

Instrumen ini memungkinkan pemilik UMKM X melakukan evaluasi mandiri dan eksperimen format konten secara berkelanjutan.

4.6 Rangkuman

Analisis data menghasilkan beberapa temuan utama terkait strategi konten

Instagram UMKM X:

1. Pemetaan Klaster ($K = 3$): Pemodelan 180 data unggahan menghasilkan tiga klaster valid (*Silhouette Coefficient* = 0,709; ANOVA $F = 1080,968$; $X^2 = 360,00$; $p = 0,000$):
 - 1) C1 (147 unggahan): Didominasi *Carousel* Sabtu pukul 17.00 bernuansa *hard-selling* penuh, dengan rerata *Engagement Rate* (ER) rendah (0,0295).
 - 2) C2 (2 unggahan): Didominasi Video/Reels Selasa–Rabu pukul 14.00 & 16.00, mencatatkan rerata ER tertinggi (0,8699) akibat respons positif visual dinamis.
 - 3) C3 (31 unggahan): Didominasi *Carousel* Minggu pukul 18.00 bernuansa *soft-selling*, menghasilkan ER stabil di tingkat menengah (0,1402).
2. Evaluasi Strategi: Frekuensi unggah tinggi tidak berbanding lurus dengan ER. Konten *hard-selling* di jam sibuk (C1) kurang efektif memicu interaksi, sedangkan format video pendek (C2) dan pesan selektif (C3) lebih sukses membangun *Attention* dan *Interest*.
3. Rekomendasi Pemasaran: UMKM X disarankan menyeimbangkan variasi format dan jadwal unggah. Video pendek diprioritaskan untuk memicu *Attention* dan *Interest*, lalu dipadukan dengan *soft-selling* (C3) guna menjaga *Desire* hingga mendorong tindakan *Action*.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan, penelitian ini menyimpulkan tiga hal utama. Pertama, pola komunikasi pemasaran Instagram UMKM X masih didominasi pendekatan *hard-selling* berformat *Carousel* yang menerapkan seluruh elemen AIDA secara transaksional penuh. Kedua, algoritma *K-Means* ($K = 3$) berhasil memetakan 180 unggahan ke dalam tiga kluster valid: Kluster 1 (147 unggahan *Carousel hard-selling*, *Engagement Rate* rendah 0,0295), Kluster 2 (2 unggahan video pendek, *Engagement Rate* tertinggi 0,8699), dan Kluster 3 (31 unggahan *Carousel soft-selling*, *Engagement Rate* menengah 0,1402). Ketiga, strategi transaksional penuh terbukti kurang efisien memicu keterlibatan audiens, sedangkan format video pendek (Reels) jauh lebih efisien dalam memicu tahapan awal AIDA, khususnya *Attention* dan *Interest*.

5.2 Saran

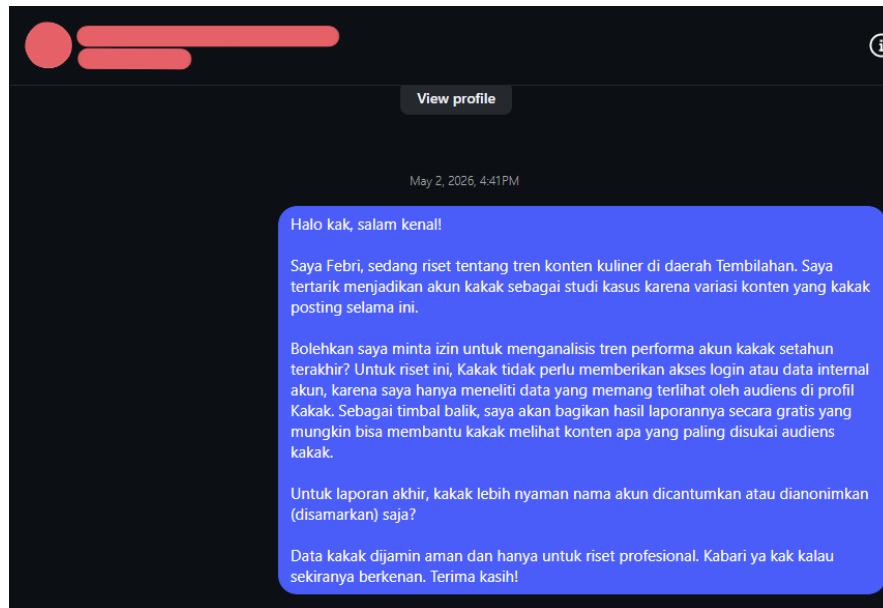
Secara praktis, manajemen UMKM X disarankan mengurangi konten *hard-selling* dan berfokus pada video pendek dinamis untuk menarik *Attention* dan *Interest*, yang lalu dipadukan dengan *soft-selling* selektif (seperti Kluster 3) guna menjaga *Desire* hingga mendorong *Action*. UMKM X juga dapat memanfaatkan prototipe Excel berbasis parameter *centroid* dan jarak *Euclidean* sebagai alat audit kinerja konten mandiri. Secara akademis, penelitian selanjutnya disarankan memperluas metrik kuantitatif (*reach*, *impressions*, *saves*, *watch time*), mengeksplorasi algoritma klasterisasi lain, serta memperpanjang rentang waktu penarikan data guna menangkap tren perilaku audiens secara lebih akurat.

DAFTAR PUSTAKA

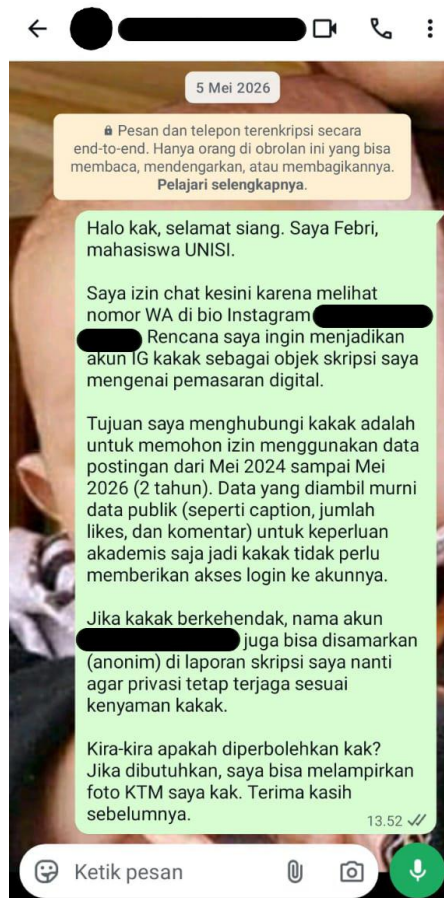
- [1] A. Nasution, "Strategi Pemasaran Bisnis Kuliner Menggunakan Influencer Melalui Media Sosial Instagram," *J. Bisnis Corp.*, vol. 6, no. 1, hal. 11–18, 2021, doi: <https://doi.org/10.46576/jbc.v6i1.1484>.
- [2] W. Salsabila, *Pengaruh Soft Selling Content Terhadap Purchase Intention Yang Dimediasi Emotional Engagement (Studi Pada Brand Ecinos Di Pulau Jawa)*. 2025. [Daring]. Tersedia pada: <http://etheses.uin-malang.ac.id/id/eprint/83214>
- [3] A. R. Yanti, "Strategi Komunikasi Pemasaran Brand Skincare Mako By Seris Dalam Meningkatkan Engagement Melalui Media Sosial," *J. Manag. Educ. Commun.*, vol. 1, no. 3, 2023, doi: <https://doi.org/10.64685/jmec.v1i2.11>.
- [4] A. Rahman, R. A. Kusumadewi, dan F. Nurzaman, "Strategi Pemasaran Digital Dalam Meningkatkan Daya Saing Perusahaan Di Era Transformasi Teknologi," *Idarah J. Manaj. Pemasar.*, vol. 2, no. 1, hal. 8–20, 2025, doi: <https://doi.org/10.70283/idarrah.v2i1.64>.
- [5] F. I. Lubis dan M. Iqbal, "Analisis Waktu Optimal Unggah Instagram Menggunakan K-Means Clustering pada Data Sintetis," vol. 5, no. 1, hal. 1497–1504, 2026, doi: <https://doi.org/10.31004/riggs.v5i1.6042>.
- [6] . Nurhidayat, N. Fadillah, Syahrul, S. Dillah, dan D. F. Surianto, "K-Means-Based Behavioral Segmentation of Social Media Users for Digital Communication Analysis in Indonesia," vol. 5, no. 2, hal. 249–261, 2025, doi: [10.24042/ijecs.v5i2.28749](https://doi.org/10.24042/ijecs.v5i2.28749).
- [7] E. W. Ambarsari, D. Fathudin, dan G. Alfiani, "Utilizing K-Means Clustering to Understanding Audience Interest in SEO- Optimized Media Content," vol. 3, no. 2, hal. 208–214, 2024, doi: [10.47065/comforch.v3i2.1207](https://doi.org/10.47065/comforch.v3i2.1207).
- [8] W. E. Kedi, C. Ejimuda, dan C. Idemudia, "Machine learning software for optimizing SME social media marketing campaigns," vol. 5, no. 7, hal. 1634–1647, 2024, doi: [10.51594/csitrj.v5i7.1349](https://doi.org/10.51594/csitrj.v5i7.1349).
- [9] A. Nallasivam, G. Desai, D. Bandhu, dan I. U. Jagadeeswari, "Predicting Consumer Adoption of Luxury Products via Instagram Marketing : A Machine Learning Approach Anitha Nallasivam Guruprasad Desai Indonesia," hal. 1–39, 2025, [Daring]. Tersedia pada: <https://f1000research.com/articles/14-890>
- [10] D. A. Nugraha, T. E. Priandono, N. F. Ainiyyah, dan Y. Ilham, "Analisis Isi Konten Media Sosial Instagram Pemerintah Indonesia di Tingkat Kementerian dan Lembaga Content Analysis of the Indonesian Government ' s Instagram Accounts at the Ministry and Agency Levels," vol. 9, no. 1, hal. 42–52, 2025, doi: <https://doi.org/10.56873/jimik.v9i1.532>.
- [11] A. S. Nafsyah, S. R. Maulidyah, A. S. Nurlia, dan W. Putri, "Analisis

- Konten Media Sosial Instagram By . U Sebagai Media Penyebaran Informasi & Komunikasi,” vol. 2, no. 2, hal. 1–11, 2022, doi: <https://doi.org/10.35842/massive.v2i2.57>.
- [12] R. Yadav, S. Tripathi, dan S. S. Shukla, “AIDA Model For Social Media Marketing,” vol. 01, no. 02, hal. 1–4, 2025, [Daring]. Tersedia pada: <https://jdacm.com/index.php/jdacm/article/view/9>
 - [13] U. Rahardja, Q. Aini, R. Supriati, dan M. A. Al Hafiz, “Optimizing Digital Promotional Graphic Design Strategies Using the AIDA Model,” vol. 7, no. 2, hal. 220–230, 2026, doi: <https://doi.org/10.34306/ajri.v7i2.1425>.
 - [14] M. Sarfraz, B. Al Kurdi, dan M. Rafiq, “How Digital Marketing Shapes Consumer Decision-Making Employing (AIDA) Model with Respect to Consumer Knowledge and Consumer Experience,” hal. 39–48, 2025, doi: <https://doi.org/10.64251/ijmmi.66>.
 - [15] N. Mirantika dan E. Rijanto, “Comparative Analysis Of K-Means And K-Medoids Algorithms In Determining Customer Segmentation Using Rfm Model,” vol. 18, no. 5, hal. 2340–2351, 2023, [Daring]. Tersedia pada: https://jestec.taylors.edu.my/Vol 18 Issue 5 October 2023/18_5_06.pdf
 - [16] B. Siregar, “Comparative Analysis of K-Means and K-Medoids Algorithms for Product Sales Clustering and Customer,” vol. 7, no. 2, hal. 360–370, 2024, [Daring]. Tersedia pada: <https://journal.unm.ac.id/index.php/JMATHCOS/article/view/4053>
 - [17] H. Arief, M. T. Sembiring, dan E. S. Rini, “Strategy To Increase Followers and Account Reach Social Media Instagram PT. Bank SUMUT,” vol. 9, no. 2, 2026, doi: [10.32734/lwsa.v9i2.2845](https://doi.org/10.32734/lwsa.v9i2.2845).

LAMPIRAN



Gambar 1. Pesan Izin Ke UMKM X melalui Instagram



Gambar 2. Pesan Izin ke UMKM X melalui WhatsApp