ANALISIS SENTIMEN MASYARAKAT INDONESIA PADA KEWAJIBAN SERTIFIKASI HALAL MENGGUNAKAN METODE *K-MEANS CLUSTER*ING

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Strata-1
Program Studi Teknik Industri Program Sarjana - Fakultas Teknologi Industri
Universitas Islam Indonesia

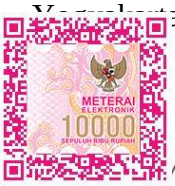
Nama : Samaya Dhiya Salindri
No. Mahasiswa : 22522107

PROGRAM STUDI TEKNIK INDUSTRI PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2026

PERNYATAAN KEASLIAN

Saya mengakui bahwa tugas akhir ini adalah hasil karya saya sendiri kecuali kutipan dan ringkasan yang seluruhnya sudah saya jelaskan sumbernya. Jika di kemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah maka saya bersedia ijazah yang telah saya terima ditarik kembali oleh Universitas Islam Indonesia.

Yogyakarta, 13 Mei 2026



(Samaya Dhiya Salindri)
22522107

SURAT BUKTI PENELITIAN

UNIVERSITAS ISLAM INDONESIA

FAKULTAS TEKNOLOGI INDUSTRI
Gedung KH. Mas Mansur
Kampus Terpadu Universitas Islam Indonesia
Jl. Kaliurang km 14,5 Yogyakarta 55584
T. (0274) 898444 ext. 4100, 4101
F. (0274) 895007
E. fti@uii.ac.id
W. fti.uii.ac.id

SURAT KETERANGAN PENELITIAN

Nomor: 013/Ka.Lab.Datmin/70/Lab.Datmin/V/2026

Assalamu'alaikum Warahmatullahi Wabarakatuh

Kami yang bertanda tangan dibawah ini, menerangkan bahwa mahasiswa dengan keterangan sebagai berikut :

Nama	:	Samaya Dhiya Salindri
No. Mhs	:	22522107
Dosen Pembimbing	:	Dr. Harwati, S.T., M.T.

Telah selesai melaksanakan penelitian yang berjudul " ANALISIS SENTIMEN MASYARAKAT INDONESIA PADA KEWAJIBAN SERTIFIKASI HALAL MENGGUNAKAN METODE K-MEANS CLUSTERING" di Laboratorium Data Mining, Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 21 Desember 2025 sampai dengan tanggal 21 Februari 2026.

Demikian surat keterangan kami keluarkan, agar dapat dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Yogyakarta, 18 Mei 2026

Kepala Laboratorium
Data Mining

Dr. Dwi Adi Purnama, S.T.

LEMBAR PENGESAHAN PEMBIMBING

ANALISIS SENTIMEN MASYARAKAT INDONESIA PADA KEWAJIBAN SERTIFIKASI HALAL MENGGUNAKAN METODE *K-MEANS CLUSTER*ING

TUGAS AKHIR

Disusun Oleh :

Nama : Samaya Dhiya Salindri

No. Mahasiswa : 22522107

Yogyakarta, 13 Mei 2026

Dosen Pembimbing

(Dr. Harwati, S.T., M.T.)

LEMBAR PENGESAHAN DOSEN PENGUJI

ANALISIS SENTIMEN MASYARAKAT INDONESIA PADA KEWAJIBAN SERTIFIKASI HALAL MENGGUNAKAN METODE *K-MEANS CLUSTER*ING

TUGAS AKHIR

Disusun Oleh :

Nama : Samaya Dhiya Salindri
No. Mahasiswa : 22 522 107

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata-1 Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 3 Juli 2026

Tim Penguji

Dr. Harwati, S.T., M.T. Ketua	
Dr. Joko Sulistio, S.T., M.Sc. Anggota I	
Wahyudhi Sutrisno, S.T., M.M. Anggota II	020726

Mengetahui,
Ketua Program Studi Teknik Industri Program Sarjana
Fakultas Teknologi Industri
Universitas Islam Indonesia

Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM
NIK. 015220101

HALAMAN PERSEMBAHAN

Rasa syukur saya panjatkan kepada Allah SWT atas pertolongan dan hidayah-Nya sehingga penelitian dapat terselesaikan. Orang tua saya tercinta Ibu Cecilia Diam Arishanty dan almarhum Bapak Teguh Indriyono atas doa, kasih sayang, dukungan, dan bimbingan yang telah diberikan selama ini. Terima kasih saya sampaikan kepada Ibu Dr. Harwati, S.T., M.T. selaku dosen pembimbing atas segala bimbingan, arahan, dan dukungan yang diberikan selama proses penelitian ini.

MOTTO

”*Maka sesungguhnya bersama kesulitan ada kemudahan. Sesungguhnya bersama kesulitan ada kemudahan*”

(Q.S. Al-Insyirah: 5-6)

”Sesungguhnya Aku dekat. Aku mengabulkan permohonan orang yang berdoa apabila ia berdoa kepada-Ku.”

(Q.S. Al-Baqarah: 186)

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, taufik, hidayah, serta nikmat-Nya sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul "Analisis Sentimen Masyarakat Indonesia Pada Kewajiban Sertifikasi Halal Menggunakan Metode *K-Means Cluster*ing" dengan baik. Shalawat serta salam senantiasa tercurah kepada Nabi Muhammad SAW yang telah membawa umat manusia dari zaman kegelapan menuju jalan yang terang.

Laporan Tugas Akhir ini disusun sebagai salah satu syarat memperoleh gelar Sarjana Strata Satu (S1) pada Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia. Tujuan penyusunan Tugas Akhir ini adalah menerapkan ilmu pengetahuan yang diperoleh selama perkuliahan. Dalam proses pelaksanaan Tugas Akhir, penulis mendapatkan banyak bantuan, dukungan, dan kesempatan dari berbagai pihak secara langsung maupun tidak langsung. Oleh karena itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Ir. Hari Purnomo, M.T., IPU., ASEAN.Eng. selaku Dekan Fakultas Teknologi Industri, Universitas Islam Indonesia.
2. Bapak Dr. Drs. Imam Djati Widodo, M.Eng.Sc. selaku Ketua Jurusan Teknik Industri, Universitas Islam Indonesia.
3. Bapak Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM. Selaku Ketua Program Studi S1 Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia.
4. Ibu Dr. Harwati, S.T., M.T. selaku dosen pembimbing laporan Tugas Akhir yang telah memberikan bimbingan, arahan, motivasi, dan ilmu sehingga penulis mampu menyelesaikan Tugas Akhir ini.
5. Kepada kedua orang tua saya tercinta, Ibu Cecilia Diam Arishanty dan mendiang ayah saya tersayang almarhum Bapak Teguh Indriyono, yang telah mengiringi setiap langkah penulis dengan doa dan kasih sayang. Kemudian, kepada saudara kandung terkasih, Nanda Putri Mintari, yang senantiasa memberikan doa, dukungan, dan semangat dalam setiap proses penulis.
6. Kepada sahabat seperjuangan Teknik Industri 2022, terutama Atiqoh Fairuz Hamidah, Qonita Silvia Faizun, dan Amanda Trixie Alsyaviar yang telah menjadi sahabat terdekat seperjuangan sejak awal perkuliahan. Kemudian, kepada Rani Novalentina, Angela Julianingsih, dan Titania Citra Ayu Rismayanti yang selalu memberi dukungan selama masa perkuliahan.
7. Kepada sahabat seperjuangan SMA, Titis Retno Hidayati dan Agban Nur Urbani yang selalu menghibur, mendukung, dan mengingatkan penulis di setiap proses penulis.
8. Kepada teman-teman Asisten Laboratorium Data Mining yang telah memberikan dukungan, arahan, hiburan selama proses perkuliahan dan pelaksanaan Tugas Akhir.
9. Seluruh pihak lain yang tidak dapat dituliskan satu persatu yang telah memberikan dukungan selama pelaksanaan Tugas Akhir.

Penulis menyadari bahwa laporan Tugas Akhir ini masih memiliki kekurangan dan belum sepenuhnya sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun untuk perbaikan di masa mendatang. Semoga laporan ini dapat bermanfaat bagi pembaca dan penelitian selanjutnya. *Wassalamu'alaikum Warahmatullahi Wabarakatuh*.

Yogyakarta, 12 Mei 2026
Samaya Dhiya Salindri
NIM. 22522107

ABSTRAK

Pasar halal global terus mengalami pertumbuhan, Indonesia memiliki posisi strategis sebagai negara dengan populasi Muslim terbesar di dunia. Untuk memperkuat jaminan produk halal, pemerintah menerapkan kebijakan kewajiban sertifikasi halal bagi produk yang beredar dan diperdagangkan di Indonesia. Akan tetapi, penerapan kebijakan ini masih menghadapi tantangan berupa penerapan kebijakan pada pelaku usaha UMKM yang belum sesuai target. Kondisi tersebut memunculkan berbagai opini masyarakat di media sosial khususnya YouTube dan Instagram sehingga diperlukan analisis untuk mengetahui kecenderungan sentimen dan pola opini publik. Berdasarkan hasil analisis opini masyarakat, akan dirumuskan strategi komunikasi dan sosialisasi agar kebijakan dapat berjalan efektif dan dapat mencapai target. Data yang digunakan berupa komentar YouTube dan Instagram pada periode April 2024 sampai April 2026. Metode penelitian yang digunakan adalah *text mining*, penentuan jumlah *cluster* optimal menggunakan metode Elbow, dan pengelompokkan data menggunakan algoritma *K-Means Clustering*. Hasil penelitian menunjukkan terbentuknya enam *cluster* opini, yaitu prosedur sertifikasi UMKM, kecurigaan terhadap komersialisasi oleh oknum, kritik kebijakan dan peran negara, aturan labeling produk nonhalal, isu korupsi dan pungutan liar oleh pemerintah, dan polemik produk nonhalal. *Cluster* dengan sentimen negatif menjadi *Cluster* optimal yang akan diberikan strategi kebijakan. Strategi yang diusulkan mencakup perbaikan komunikasi manfaat, penguatan peran pemerintah sebagai fasilitator, penyederhanaan prosedur, pendampingan UMKM, dan transparansi biaya.

Kata Kunci: Analisis Sentimen, Sertifikasi Halal, *K-Means Cluster*ing, Media Sosial

DAFTAR ISI

PERNYATAAN KEASLIAN ii
SURAT BUKTI PENELITIAN iii
LEMBAR PENGESAHAN PEMBIMBING iv
LEMBAR PENGESAHAN DOSEN PENGUJI v
HALAMAN PERSEMBAHAN vi
MOTTO vii
KATA PENGANTAR viii
ABSTRAK x
DAFTAR ISI xi
DAFTAR TABEL xiii
DAFTAR GAMBAR xiv
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Tujuan Penelitian 5
1.4 Manfaat Penelitian 6
1.5 Batasan Penelitian 6
1.6 Sistematika Penulisan 7
BAB II TINJAUAN PUSTAKA 8
2.1 Kajian Literatur 8
2.2 Landasan Teori 16
2.2.1 Analisis Sentimen 16
2.2.2 Text Mining 16
2.2.3 Text Preprocessing 16
2.2.4 Term Frequency-Inverse Document Frequency (TF-IDF) 17
2.2.5 Supervised Learning 18
2.2.6 Unsupervised Learning 18
2.2.7 *K-Means* Clustering 18
2.2.8 Tahapan *K-Means* Clustering 18
2.2.9 Penentuan Jumlah Cluster dengan Metode Elbow 19
2.2.10 Aplikasi Instagram 19
2.2.11 Aplikasi YouTube 20
BAB III METODE PENELITIAN 21
3.1 Objek Penelitian 21
3.2 Jenis Data 21
3.3 Metode Pengumpulan Data 21
3.4 Alur Penelitian 22
3.5 Perhitungan dalam Pembobotan TF-IDF 27
3.6 Perhitungan dalam Penentuan Jumlah *Cluster* dengan Elbow 27
3.7 Perhitungan *K-Means Cluster*ing 27
BAB IV PENGUMPULAN DAN PENGOLAHAN DATA 29
4.1 Pengumpulan Data 29
4.1.1 Aplikasi YouTube dan Instagram 29
4.2 Pengolahan Data 32

4.2.1 Pre-Processing Data 32
4.2.2 Representasi Teks Menggunakan TF-IDF 34
4.2.3 Penentuan Jumlah Cluster dengan Metode Elbow 35
4.2.4 Pengelompokkan Data dengan *K-Means* Clustering 36
4.2.5 Profilisasi dan Labeling Cluster 37
4.2.6 Distribusi Sentimen 42
BAB V PEMBAHASAN 44
5.1 Deskripsi *Cluster* 44
5.2 Strategi *Cluster* 48
5.2.1 Strategi terhadap Cluster Kecurigaan Terhadap Oknum dan Komersialisasi 50
5.2.2 Strategi terhadap Cluster Kritik Kebijakan dan Peran Negara 51
5.2.3 Strategi terhadap Cluster Dugaan Pungutan dan Korupsi 53
5.3 Limitasi Penelitian 56
BAB VI PENUTUP 58
6.1 Kesimpulan 58
6.2 Saran 59
DAFTAR PUSTAKA 60
LAMPIRAN A-1

DAFTAR TABEL

Tabel 2. 1 Posisi Penelitian 11
Tabel 4. 1 Sampel Data *Scraping*.....31
Tabel 4. 2 *Data Cleaning* 32
Tabel 4. 3 *Case Folding* 32
Tabel 4. 4 Normalisasi Slang 33
Tabel 4. 5 Tokenisasi 33
Tabel 4. 6 Penghapusan *Stopword* 33
Tabel 4. 7 Proses *Stemming*.....34
Tabel 4. 8 Pembobotan TF-IDF 34
Tabel 4. 9 Selisih/Gap Nilai WCSS Pembentukan Jumlah *Cluster* 35
Tabel 4. 10 Profilisasi dan Labeling *Cluster*.....41
Tabel 5. 1 Deskripsi *Cluster*.....45
Tabel 5. 2 Strategi *Cluster* Optimal 49

DAFTAR GAMBAR

Gambar 3. 1 Diagram Alur Penelitian Bagian 1 22
Gambar 3. 2 Diagram Alur Penelitian Bagian 2 23
Gambar 4. 1 Grafik Metode *Elbow* 36
Gambar 4. 2 Distribusi Jumlah Komentar Setiap *Cluster* 37
Gambar 4. 3 Grafik Kata Dominan *Cluster* 0 38
Gambar 4. 4 Grafik Kata Dominan *Cluster* 1 38
Gambar 4. 5 Grafik Kata Dominan *Cluster* 2 39
Gambar 4. 6 Grafik Kata Dominan *Cluster* 3 39
Gambar 4. 7 Grafik Kata Dominan *Cluster* 4 40
Gambar 4. 8 Grafik Kata Dominan *Cluster* 5 40
Gambar 4. 9 Grafik Distribusi Sentimen.....43
Gambar 5. 1 Persentase Distribusi Sentimen 44
Gambar 5. 2 *Design Engineering* Dashboard Transparansi Akuntabilitas BPJPH.....55

BAB I
PENDAHULUAN

1.1 Latar Belakang

Pasar halal global menunjukkan pertumbuhan pesat selama beberapa dekade terakhir. Laporan *State of the Global Islamic Economy* menunjukkan bahwa belanja konsumen pada sektor halal pada tahun 2023 mencapai sekitar 2,43 triliun dolar AS dan diperkirakan meningkat menjadi 3,36 triliun dolar AS pada tahun 2028 dengan sektor makanan halal sebagai kontributor terbesar (Kawsar, 2025). Pada dinamika pasar tersebut, Indonesia memiliki posisi strategis dalam pengembangan pasar halal global sebagai negara dengan populasi Muslim terbesar di dunia, yaitu sekitar 251 juta jiwa atau 87,15 persen dari total penduduknya (Kementerian Dalam Negeri, 2026) sehingga kebutuhan terhadap jaminan produk halal menjadi bagian penting dalam perlindungan konsumen, peningkatan kepercayaan masyarakat, dan penguatan daya saing produk nasional.

Dalam upaya memperkuat industri halal nasional, pemerintah menetapkan pemberlakuan kebijakan kewajiban sertifikasi halal di Indonesia melalui Undang-Undang Nomor 33 Tahun 2014 tentang Jaminan Produk Halal dan Undang-Undang Nomor 6 Tahun 2023 yang diatur lebih lanjut dalam Peraturan Pemerintah Nomor 42 Tahun 2024 mengenai penyelenggaraan bidang jaminan produk halal. Kebijakan ini mewajibkan produk makanan, minuman, bahan baku, obat-obatan, kosmetik, dan jasa penyembelihan yang masuk, beredar, dan diperdagangkan di Indonesia wajib memiliki sertifikat halal kecuali produk dari bahan haram yang wajib diberi keterangan tidak halal. Pemberlakuan kewajiban ini mulai diberlakukan pada Oktober 2024 untuk pelaku usaha menengah dan besar. Sedangkan, bagi pelaku Usaha Mikro dan Kecil (UMK), terutama produk makanan dan minuman diperpanjang sampai 2026 (Wiryono, 2024).

Berdasarkan data Kementerian UMKM, jumlah UMKM di Indonesia mencapai sekitar 30,18 juta yang sudah terdata dalam sistem informasi data tunggal pemerintah (Indonesian Chamber of Commerce and Industry, 2024). Untuk mendukung segera terlaksananya sertifikasi halal, BPJPH membuat program SEHATI atau Sertifikasi Halal Gratis bagi pelaku UMK dengan kuota sebesar 1 juta sertifikat halal gratis pada tahun 2025 dan 1,35 juta sertifikat halal gratis pada tahun 2026 (Badan Penyelenggara Jaminan Produk Halal, 2026). Namun, jumlah

UMKM yang terdaftar sertifikasi halal masih belum optimal. Data Badan Penyelenggara Jaminan Produk Halal (BPJPH) menunjukkan bahwa dari 28 juta UMKM yang menjadi target sertifikasi halal, saat ini baru sekitar 3 juta UMKM yang telah tersertifikasi halal yang setara dengan 10,71% (CNBC Indonesia, 2026). Padahal pemerintah telah menargetkan produk makanan dan minuman UMKM yang beredar wajib bersertifikasi halal paling lambat Oktober 2026. Hal ini menunjukkan bahwa tingkat penerapan kebijakan masih rendah dibandingkan dengan target dan jumlah pelaku usaha yang ada.

Rendahnya tingkat penerapan kebijakan ini dipengaruhi oleh berbagai tantangan. Beberapa di antaranya seperti keterbatasan biaya sertifikasi halal bagi pelaku usaha kecil, prosedur administrasi yang kompleks, serta keterbatasan akses informasi mengenai sertifikasi halal. Penelitian (Sholikhah et al., 2025) menunjukkan bahwa literasi halal dan persepsi terhadap biaya sertifikasi halal memiliki pengaruh signifikan terhadap minat pelaku usaha untuk memiliki sertifikat halal. Selain itu, kurangnya sosialisasi dan pemahaman masyarakat terhadap prosedur sertifikasi juga menjadi hambatan utama dalam implementasi kebijakan (Hikmah et al., 2025). Hal ini membuat banyak pelaku usaha belum menjadikan sertifikasi halal sebagai prioritas. Sebagian masyarakat menganggapnya sebagai beban tambahan pada operasional usaha (Sandi, 2026). Dengan demikian, penerimaan masyarakat terhadap kewajiban sertifikasi halal belum merata dan dipengaruhi faktor ekonomi, pemahaman serta persepsi terhadap regulasi.

Kondisi ini menunjukkan bahwa kewajiban sertifikasi halal tidak hanya berkaitan dengan kesiapan regulasi dan administrasi tetapi juga penerimaan masyarakat terhadap kebijakan ini. Pada praktiknya, penerapan kewajiban sertifikasi halal menimbulkan beragam tanggapan dari masyarakat. Sebagian masyarakat mendukung kebijakan ini karena dianggap dapat memberikan jaminan keamanan dan memperkuat kualitas produk yang ada di pasar (Pradana et al., 2024). Akan tetapi masih terdapat banyak masyarakat dan pelaku usaha yang memandang kebijakan ini sebagai beban tambahan terutama untuk pelaku usaha kecil yang terbatas dalam memahami aturan sertifikasi halal, biaya, serta pendampingan dalam proses pembuatan sertifikasi halal yang membutuhkan banyak waktu dan usaha terutama untuk mengumpulkan dokumen persyaratan dan pelaksanaan audit halal (Prawiro & Fathudin, 2023). Perbedaan pendapat ini menunjukkan bahwa penerapan kebijakan sertifikasi halal tidak hanya membutuhkan persiapan dari lembaga penyelenggara tetapi juga pemahaman mengenai

persepsi masyarakat sebagai pihak yang terdampak langsung ataupun tidak langsung oleh kebijakan ini (Denha et al., 2024).

Meskipun penelitian mengenai sertifikasi halal telah banyak membahas aspek regulasi, kesiapan pelaku UMK, literasi halal, biaya, dan perlindungan konsumen. Akan tetapi, penelitian yang memetakan persepsi masyarakat Indonesia mengenai kewajiban sertifikasi halal melalui komentar di *platform* media sosial masih terbatas. Media sosial telah memberikan ruang untuk masyarakat menyampaikan pendapat, kritik, dukungan, maupun pengalaman pribadi terkait kebijakan tertentu (Ariel et al., 2022). Permasalahan sertifikasi halal tidak hanya dibahas dalam forum resmi atau pemberitaan tetapi juga melalui komentar masyarakat di berbagai *platform* digital. Berbagai jenis komentar tersebut perlu dikaji lebih dalam agar dapat memberikan gambaran mengenai tingkat penerimaan masyarakat terhadap kebijakan kewajiban sertifikasi halal.

Media sosial telah menjadi sarana utama masyarakat dalam menyampaikan opini terkait permasalahan publik. Berbagai *platform* seperti X (Twitter), Facebook, TikTok, YouTube, dan Instagram dapat menjadi sumber data dalam mengamati opini masyarakat. Akan tetapi, setiap *platform* memiliki karakteristik yang berbeda. X umumnya berisi opini terkait isu yang sedang berkembang pesat berbentuk teks singkat yang bersifat *microblogging* sehingga terdapat tantangan pencarian data teks dengan konteks yang terbatas (Murthy, 2024). TikTok lebih berfokus pada konten video durasi pendek sehingga data tekstual komentar yang dihasilkan relatif terbatas dibandingkan *platform* lainnya (Zannettou et al., 2024). Kemudian, Facebook memiliki berbagai bentuk interaksi melalui halaman maupun komunitas tetapi aktivitas pengguna berlangsung pada ruang yang memiliki pengaturan privasi tertentu dimana data publik untuk kepentingan penelitian tidak dapat diakses secara bebas (Kim et al., 2021).

Sementara itu, *platform* YouTube dan Instagram telah banyak digunakan masyarakat Indonesia untuk mendapatkan informasi dan menyampaikan pendapat terhadap berbagai permasalahan publik. Pada awal 2025, terdapat 143 juta pengguna YouTube dan 103 juta pengguna Instagram (DataReportal, 2025). Kedua *platform* tersebut juga memiliki karakteristik data yang berbeda tetapi saling melengkapi dalam penelitian media sosial. Instagram sebagai *platform* berbasis visual menghasilkan data dari *caption*, *hashtag*, dan komentar yang bersifat spontan dan singkat sedangkan YouTube sebagai *platform* video memungkinkan interaksi lebih kompleks dengan diskusi lebih panjang (Quan-Haase & Sloan, 2022). Selain itu, pada kedua *platform* tersebut BPJPH aktif menggunakannya sebagai media komunikasi dan sosialisasi

sertifikasi halal pada masyarakat terutama akun instagram resmi BPJPH yang telah diteliti sebagai media efektif dalam penyebaran informasi sertifikat halal sedangkan YouTube digunakan untuk menyampaikan webinar, edukasi, maupun *podcast* terkait kebijakan sertifikasi halal (Djamaluddin et al., 2024). Dengan demikian, kedua *platform* ini dapat memberikan sudut pandang yang lebih luas mengenai persepsi publik terhadap kebijakan kewajiban sertifikasi halal. Akan tetapi, jumlah komentar yang tersebar di media sosial begitu banyak sehingga proses pengamatan opini masyarakat secara manual tidak efektif. Data komentar dengan jumlah besar membutuhkan pendekatan yang dapat mengelompokkan dan memahami kecenderungan opini secara sistematis. Oleh karena itu, analisis terhadap sentimen masyarakat di media sosial terkait kewajiban sertifikasi halal penting dilakukan.

Berdasarkan permasalahan tersebut, penelitian ini berfokus pada analisis sentimen masyarakat Indonesia terhadap kewajiban sertifikasi halal melalui komentar pada *platform* YouTube dan Instagram. Sentimen merupakan opini dan pandangan berdasarkan perasaan yang berlebihan terhadap suatu hal sehingga sentimen digunakan untuk mengetahui perasaan seseorang terhadap suatu topik atau objek (Fernandes et al., 2023). Analisis sentimen merupakan metode yang dipakai untuk mengkaji data opini dan mengolah data teks secara otomatis untuk mengetahui sentimen yang terdapat di dalamnya. Penelitian ini menggunakan pendekatan *K-Means Clustering* dengan bantuan metode *Elbow* dalam mengelompokkan komentar masyarakat ke dalam *cluster* optimal berdasarkan pola teks. Pemilihan pendekatan tersebut diharapkan dapat menggambarkan kecenderungan opini masyarakat dengan lebih terstruktur karena *clustering* dapat digunakan untuk mengelompokkan data teks yang tidak terstruktur (Muhariya et al., 2023) dan Elbow digunakan untuk membantu menentukan jumlah *cluster* optimal (Wicaksana et al., 2023). Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam memahami dinamika persepsi masyarakat secara menyeluruh dan menjadi pertimbangan bagi pemerintah dalam menciptakan strategi komunikasi dan sosialisasi kebijakan yang lebih efektif.

1.2 Rumusan Masalah

Kewajiban sertifikasi halal pada dasarnya diterapkan sebagai upaya untuk memberikan jaminan kehalalan produk, memperkuat perlindungan konsumen, dan meningkatkan daya saing usaha terutama untuk pelaku UMK. Namun, realitanya kebijakan ini masih menghadapi beberapa hambatan seperti keterbatasan biaya, proses administrasi, akses informasi, serta pemahaman

masyarakat yang belum merata mengenai manfaat sertifikasi halal. Kondisi ini mendorong munculnya berbagai opini masyarakat di media sosial terutama pada *platform* YouTube dan Instagram. Jika opini tersebut tidak dianalisis secara sistematis maka kebijakan sertifikasi halal berpotensi mengalami hambatan dalam penerimaan publik, resistensi dari pelaku usaha, dan strategi sosialisasi yang tidak tepat sasaran. Berdasarkan hal tersebut, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kecenderungan sentimen masyarakat Indonesia terhadap kebijakan kewajiban sertifikasi halal pada komentar YouTube dan Instagram?
2. Bagaimana pola opini masyarakat yang muncul terhadap kebijakan kewajiban sertifikasi halal pada komentar YouTube dan Instagram?
3. Bagaimana hasil analisis opini masyarakat dapat membantu dalam strategi komunikasi dan sosialisasi kebijakan strategi halal di Indonesia?

1.3 Tujuan Penelitian

Penelitian yang dilakukan memiliki tujuan sebagai berikut:

1. Mengetahui kecenderungan sentimen masyarakat Indonesia terhadap kebijakan kewajiban sertifikasi halal pada komentar YouTube dan Instagram.
2. Mengetahui pola opini masyarakat yang muncul terhadap kebijakan kewajiban sertifikasi halal pada komentar YouTube dan Instagram.
3. Merumuskan strategi komunikasi dan sosialisasi kebijakan strategi halal di Indonesia berdasarkan hasil analisis opini masyarakat.

1.4 Manfaat Penelitian

Penelitian yang dilakukan memiliki manfaat sebagai berikut:

1. Bagi Akademik

   Penelitian ini memberikan kontribusi pengetahuan pada kajian analisis sentimen kebijakan publik melalui media sosial dengan metode *K-Means Clustering*. Hasil penelitian dapat menjadi referensi untuk pengembangan pada penelitian selanjutnya.

2. Bagi Pemerintah

   Penelitian ini memberikan gambaran sentimen masyarakat dan permasalahan utama masyarakat terhadap kebijakan kewajiban sertifikasi halal. Hasil penelitian ini dapat dimanfaatkan sebagai bahan evaluasi dan perumusan strategi sosialisasi untuk menyempurnakan penerapan kebijakan agar lebih efektif.

3. Bagi Pelaku Usaha

   Penelitian ini memberikan penjelasan mengenai hasil identifikasi isu, kendala, dan kekhawatiran utama yang dimiliki oleh para pelaku usaha terkait dengan kewajiban sertifikasi halal. Hasil penelitian ini diharapkan dapat meningkatkan pemahaman pelaku usaha terhadap pentingnya sertifikasi halal.

1.5 Batasan Penelitian

Batasan penelitian yang digunakan adalah sebagai berikut:

1. Penelitian ini hanya menggunakan data komentar masyarakat pada platform YouTube dan Instagram terkait kewajiban sertifikasi halal di Indonesia dalam periode April 2024 sampai April 2026 sehingga hasil penelitian bergantung pada kualitas, kelengkapan, dan representativitas komentar yang berhasil dikumpulkan saja.
2. Penelitian ini hanya menggunakan sumber data pada komentar berbentuk teks saja, tidak menganalisis unsur lain seperti gambar, video, emoji, jumlah likes, jumlah pengikut akun.
3. Penelitian ini tidak menganalisis identitas pribadi pengguna seperti usia, jenis kelamin, agama, domisili, pekerjaan, dan latar belakang sosial ekonomi karena informasi tersebut tidak tersedia dalam data komentar yang dikumpulkan.
4. Penelitian ini hanya sampai pada tahap perumusan rekomendasi strategi berdasarkan hasil analisis sentimen dan pengelompokkan opini kebijakan sertifikasi halal di Indonesia, tanpa implementasi dan pengujian langsung strategi yang diusulkan.

1.6 Sistematika Penulisan

Berikut merupakan sistematika penulisan pada penelitian ini:

BAB I PENDAHULUAN

Bab ini berisi latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan atau ruang lingkup penelitian, manfaat penelitian, dan sistematika penulisan laporan Tugas Akhir.

BAB II KAJIAN PUSTAKA

Bab ini menjelaskan berbagai teori yang berkaitan dengan topik penelitian menggunakan referensi seperti buku, jurnal, dan penelitian terdahulu sebagai dasar untuk menyelesaikan permasalahan dalam penelitian ini.

BAB III METODE PENELITIAN

Bagian ini menjelaskan informasi mengenai objek yang diteliti, jenis data yang digunakan, metode pengumpulan data, serta diagram alur penelitian yang mencakup tahapan penelitian dari penentuan topik sampai kesimpulan dan saran.

BAB IV PENGUMPULAN DAN PENGOLAHAN DATA

Bab ini membahas proses pengumpulan dan pengolahan data menggunakan metode terpilih untuk menjawab pertanyaan penelitian pada rumusan masalah.

BAB V PEMBAHASAN

Bagian ini membahas hasil pengolahan data dari bab sebelumnya. Solusi dan rekomendasi diberikan pada pembahasan untuk menyelesaikan permasalahan penelitian.

BAB VI PENUTUP

Bab terakhir yang berisi kesimpulan mencakup hasil pembahasan yang dikaji berdasarkan tujuan penelitian. Saran juga diberikan sebagai masukan bagi pemangku kepentingan dan penelitian selanjutnya.

DAFTAR PUSTAKA

Bagian ini berisi berbagai sumber yang dipakai sebagai dasar teori dalam penelitian.

LAMPIRAN

Bagian ini memuat berbagai dokumen pendukung atau bukti yang menunjukkan terlaksananya penelitian.

BAB II
TINJAUAN PUSTAKA

2.1 Kajian Literatur

Pada penelitian ini, kajian literatur disusun berdasarkan pencarian artikel ilmiah sesuai dengan topik sertifikasi halal, opini masyarakat di media sosial, analisis sentimen, *text mining*, *K-Means Clustering*, dan metode *Elbow*. Setiap artikel diperoleh dari sumber ilmiah seperti Google Scholar, Google Books, SINTA, Garuda, ScienceDirect, dan jurnal bereputasi lainnya. Literatur yang digunakan difokuskan pada objek penelitian, metode yang digunakan, sumber data, serta hasil dan kontribusinya terhadap penelitian ini. Kajian literatur digunakan untuk memahami penelitian terdahulu dan menentukan posisi penelitian (Yam, 2024). Penelitian terdahulu berfungsi sebagai bahan untuk mengetahui perbedaan penelitian yang sedang dilakukan serta memberikan dasar dalam pemilihan metode penelitian sehingga dapat diketahui posisi penelitian ini. Berikut merupakan analisis perbedaan beberapa penelitian terdahulu terhadap penelitian yang sedang dilakukan.

Analisis Sentimen telah digunakan untuk mengkaji opini di berbagai *platform* media sosial seperti X, YouTube, Instagram, TikTok serta memetakan ulasan pada Google Play Store dan Google Maps (Setiawan, 2024). Penelitian (Ibrahim et al., 2021) juga menggunakan data dari twitter tetapi dikombinasikan dengan data dari instagram, kedua data tersebut dikaji menggunakan pendekatan *text mining* untuk mengetahui analisis sentimen pada opini publik mengenai layanan Perpustakaan Nasional yang dikategorikan menjadi sentimen positif, negatif, dan netral. Sementara itu, (Aji Nursalim & Rice Novita, 2023) memetakan ulasan dan opini publik terkait aplikasi MyPertamina dengan data dari tiga media sosial, yaitu Google Play Store, Twitter, dan YouTube menggunakan algoritma SVM yang menghasilkan sentimen positif dan negatif serta menunjukkan hasil akurasi opini YouTube tertinggi sebesar 99%. Secara umum, penelitian dalam bidang ini didominasi oleh pendekatan *supervised learning* yang digunakan untuk mengklasifikasikan sentimen ke dalam label positif, negatif, dan netral.

Selain pada berbagai *platform*, analisis sentimen diterapkan pada beragam domain ulasan layanan seperti pariwisata, aplikasi digital, dan *e-commerce*. Penelitian (Jardim et al., 2022) menggunakan analisis sentimen dengan *K-Means Clustering* dan *Elbow* untuk mengelompokkan pengguna berdasarkan sentimen dari usia dan preferensi tempat wisata di *platform* pariwisata digital. Pendekatan serupa terdapat pada penelitian (Bimantoro, 2025) *K-*

Means Clustering digunakan untuk mengelompokkan ulasan wisatawan melalui twitter pada destinasi wisata gili di Kabupaten Lombok Utara menjadi 3 *cluster* dan visualisasi topik yang sering dibahas dengan *wordcloud*. Kemudian, penelitian (Simmonds & Chowdhury, 2022) membahas analisis sentimen dengan *K-Means Clustering* yang digunakan untuk mengetahui sentimen ulasan publik pada *platform* Hulu dan Prime Video. Berikutnya, penelitian (Lestari & Febrianti, 2025) mengkaji komentar konsumen mengenai produk Shopee yang dipromosikan melalui Instagram, hasilnya dikategorikan menjadi positif dan negatif menggunakan algoritma *K-Nearest Neighbors* (KNN).

Dalam konteks kebijakan pemerintahan, analisis sentimen juga digunakan untuk menganalisis opini publik terhadap berbagai kebijakan pemerintahan seperti kenaikan pajak hiburan, pemindahan ibu kota, serta kebijakan energi dan pembatasan sosial. Penelitian (Romadhona & Isnain, 2024) mengkaji opini masyarakat pada *platform* twitter mengenai kebijakan pemerintah berupa kenaikan pajak hiburan sesuai UU No 1 Tahun 2022 yang dikategorikan menjadi positif dan negatif menggunakan algoritma SVM. Pada penelitian (Melati & Islami, 2024) analisis sentimen digunakan untuk mengidentifikasi opini masyarakat mengenai kebijakan pemindahan ibu kota menggunakan metode *K-Means Clustering*. Kemudian, terdapat penelitian (Nurfirdaus et al., 2026) yang mengkaji opini publik mengenai kebijakan pemerintah terkait campuran bioetanol 10% pada Bahan Bakar Minyak (BBM) di komentar TikTok yang dikategorikan menjadi tiga sentimen, yaitu positif, negatif, dan netral menggunakan perbandingan algoritma SVM dan *Random Forest.* Berikutnya, penelitian (Tarigan, 2026) juga mengkaji opini publik, dimana penelitian ini menggunakan data komentar dari *platform* YouTube dan X mengenai kebijakan pemerintah pada Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) yang diterapkan pada tahun 2021 untuk mengetahui topik apa yang paling sering dibicarakan dengan metode *K-Means Clustering* dan visualisasi *Word Cloud*. Sebagian besar penelitian tersebut masih menggunakan pendekatan *supervised learning* meskipun terdapat beberapa penelitian yang menggunakan *K-Means Clustering* untuk mengidentifikasi pola opini publik.

Selain menganalisis opini pada kebijakan pemerintah, analisis sentimen juga digunakan untuk mengkaji opini masyarakat terhadap produk halal dan sertifikasi halal. Penelitian (Kabir, 2022) berisi analisis sentimen opini publik terhadap koreksi misinformasi tokoh agama Muslim mengenai kehalalan vaksin COVID-19 pada komentar YouTube yang dikategorikan menjadi sentimen positif dan negatif menggunakan algoritma SVM, *naive bayes*, dan *random forest.*

Berikutnya, penelitian (Setik et al., 2021) mengkaji pendapat masyarakat mengenai sertifikasi halal di Malaysia pada *platform* twitter yang dikategorikan menjadi sentimen positif, negatif, dan netral menggunakan perbandingan metode SVM dan *naive bayes*. Pada penelitian (Afandi et al., 2023) analisis sentimen digunakan untuk mengkaji opini masyarakat Indonesia di *platform* twitter mengenai logo halal baru yang telah disahkan oleh Badan Penyelenggara Jaminan Produk Halal (BPJPH) menggunakan metode SVM.

Tidak hanya metode *supervised learning*, analisis sentimen juga dapat diterapkan menggunakan metode *unsupervised learning* seperti *K-Means Clustering* untuk mengelompokkan opini publik berdasarkan pola tertentu tanpa pelabelan. Pada penelitian (Salloum et al., 2024) menggunakan data tweet dari aplikasi twitter dengan berbagai topik bahasan yang dikategorikan ke dalam 13 *cluster* berisi berbagai emosi dengan metode *K-Means Clustering* yang dikombinasi dengan TF-IDF dan *Principal Component Analysis* (PCA) untuk visualisasi. Penelitian (Kaushik & Bhatia, 2022) juga mengkaji analisis sentimen menggunakan kombinasi metode *K-Means Clustering* dan *hierarchical clustering* untuk menganalisis topik bahasan yang sering dibicarakan oleh publik selama pandemi COVID-19 di *platform* twitter. Selanjutnya, penelitian (Harahap & Ikhsan, 2024) menggunakan metode *K-Means Clustering* dan pembobotan TF-IDF pada *pre-processing* untuk mengetahui opini publik mengenai pemilihan presiden tahun 2024 yang dikategorikan menjadi *cluster* 0 (positif), *cluster* 1 (netral), dan *cluster* 2 (negatif). Berikut merupakan tabel perbandingan penelitian terdahulu dengan penelitian ini:

Tabel 2. 1 Posisi Penelitian

No	Penelitian	Objek				Sumber Data					Pendekatan		Metode					Hasil
		Ulasan	Pariwisata	Kebijakan	Halal	X	Instagram	YouTube	TikTok	Google *Platform*	*Supervised*	*Unsupervised*	SVM	KNN	*Naive Bayes*	*Random Forest*	*K-Means Clustering*	
1	(Ibrahim et al., 2021)	✓				✓	✓											Klasifikasi sentimen 3 kategori
2	(Aji Nursalim & Rice Novita, 2023)	✓				✓		✓		✓	✓		✓					Klasifikasi sentimen 2 kategori
3	(Jardim et al., 2022)		✓							✓		✓					✓	Terbagi 3 *cluster* berdasarkan sentimen dan usia
4	(Bimantoro, 2025)		✓			✓						✓					✓	Terbentuk 3 *cluster* dan visualisasi topik yang sering muncul
5	(Simmonds & Chowdhury, 2022)	✓								✓		✓					✓	Terbentuk 2 *cluster* berdasarkan ulasan dan sentimen
6	(Lestari & Febrianti, 2025)	✓					✓				✓			✓				Klasifikasi sentimen 2 kategori

No	Penelitian	Objek			Sumber Data					Pendekatan		Metode					Hasil
		Ulasan	Pariwisata	Kebijakan	Halal	X	Instagram	YouTube	TikTok	Google *Platform*	*Supervised*	*Unsupervised*	SVM	KNN	*Naive Bayes*	*Random Forest*	*K-Means Clustering*	
7	(Romadhona & Isnain, 2024)			✓		✓					✓		✓					Klasifikasi sentimen 2 kategori
8	(Melati & Islami, 2024)			✓		✓						✓					✓	Terbentuk 3 *cluster* dengan 2 sentimen
9	(Nurfirdaus et al., 2026)			✓					✓			✓	✓			✓		Klasifikasi sentimen 3 kategori
10	(Tarigan, 2026)			✓		✓		✓				✓					✓	Terbentuk 12 *cluster* topik terkait kebijakan PPKM
11	(Kabir, 2022)				✓			✓			✓		✓		✓	✓		Klasifikasi sentimen 2 kategori
12	(Setik et al., 2021)				✓	✓					✓		✓		✓			Klasifikasi sentimen 3 kategori
13	(Afandi et al., 2023)			✓	✓	✓					✓		✓					Klasifikasi sentimen 3 kategori
14	(Salloum et al., 2024)	✓				✓						✓					✓	Terbentuk 5 *cluster* emosi utama dari 13 kategori

No	Penelitian	Objek			Sumber Data					Pendekatan		Metode					Hasil
		Ulasan	Pariwisata	Kebijakan	Halal	X	Instagram	YouTube	TikTok	Google *Platform*	*Supervised*	*Unsupervised*	SVM	KNN	*Naive Bayes*	*Random Forest*	*K-Means Clustering*	
15	(Kaushik & Bhatia, 2022)			✓		✓						✓					✓	Terbentuk 6 *cluster* emosi
16	(Harahap & Ikhsan, 2024)			✓		✓						✓					✓	Terbentuk 3 *cluster* sentimen berdasarkan pola kata
17	Penelitian ini			✓	✓		✓	✓				✓					✓	Terbentuk beberapa *cluster* berdasarkan sentimen dan karakteristik pola opini yang ada.

Berdasarkan tabel posisi penelitian di atas, analisis sentimen telah banyak digunakan pada berbagai objek dan sumber data seperti twitter (X), YouTube, Instagram, dan TikTok. Perbedaan antar penelitian dengan analisis sentimen tersebut juga didasarkan pada penggunaan jenis machine learning. Pada penelitian menggunakan *supervised learning*, terdapat hubungan antara data yang masuk (*input*) dengan data yang keluar (*output*) berdasarkan data *training* atau data latih yang telah diberi label sebelumnya secara manual seperti positif, negatif, dan netral. Model dilatih untuk mengenali pola sehingga dapat mengklasifikasikan data baru dengan tingkat akurasi yang tinggi. *Output* utamanya berupa model klasifikasi yang dapat memprediksi sentimen suatu teks atau data masukan berdasarkan data *training* atau data latihan (Birjali et al., 2021). Keunggulannya dapat menghasilkan akurasi klasifikasi tinggi apabila data *training* cukup dan representatif. Akan tetapi, *supervised learning* memiliki keterbatasan yaitu membutuhkan dataset berjumlah besar agar *output* akurasi prediksinya semakin tinggi. Selain itu, model bergantung pada domain data *training* sehingga untuk analisis sentimen yang bersifat dinamis ketika diterapkan dengan domain data yang berbeda maka performa model akurasinya dapat menurun (Lee et al., 2020). Sementara itu, *unsupervised learning* menggunakan data yang belum memiliki label sehingga algoritma akan menemukan pola, hubungan, dan struktur yang tersembunyi secara otomatis berdasarkan tingkat kemiripan antar data. Pada analisis sentimen, pendekatan *unsupervised learning* digunakan untuk mengelompokkan data yang memiliki karakteristik mirip tanpa memerlukan proses pelabelan manual. Keunggulan utamanya adalah kemampuan mengolah data dalam jumlah besar tanpa memerlukan peroses pelabelan sehingga lebih efisien dalam waktu (Bordoloi & Biswas, 2023). Pendekatan ini juga dapat diterapkan pada data yang bersifat dinamis seperti komentar media sosial karena tidak bergantung pada dataset training atau data latihan. Pada salah satu metode *unsupervised*, yatu *clustering* dapat menemukan pola tersembunyi sehingga cocok untuk penelitian yang mengkaji dan menganalisis persepsi masyarakat terhadap suatu isu (AL-Sharuee et al., 2018).

Pada penelitian ini, penggunaan *unsupervised learning* yaitu K-Means Clustering dipilih karena tujuan penelitian bukan membangun model prediksi sentimen terhadap data baru tetapi lebih berfokus pada mengidentifikasi dan mengelompokkan pola sentimen masyarakat mengenai kebijakan sertifikasi halal berdasarkan kemiripan karakteristik komentarnya. Selain itu, penelitian mengenai isu halal sebagian besar masih berfokus pada produk atau label halal dan belum secara spesifik mengkaji kebijakan sertifikasi halal. Penggunaan YouTube dan Instagram sebagai sumber data akan memberikan sudut pandang yang lebih luas dibandingkan

penelitian sebelumnya. Hal ini karena YouTube dan Instagram memiliki karakteristik data yang berbeda tetapi saling melengkapi dalam penelitian media sosial. Instagram sebagai *platform* berbasis visual menghasilkan data dari *caption*, *hashtag*, dan komentar yang bersifat spontan dan singkat sedangkan YouTube sebagai *platform* video memungkinkan interaksi lebih kompleks dengan diskusi lebih panjang. Dengan demikian, penelitian ini tidak hanya menambah literatur pada objek dan sumber data tetapi juga memberikan kontribusi metode dalam penggunaan *unsupervised learning* pada analisis sentimen kebijakan publik terhadap kewajiban sertifikasi halal di Indonesia.

2.2 Landasan Teori

2.2.1 Analisis Sentimen

Analisis sentimen merupakan proses yang menganalisis kecenderungan pandangan atau emosi yang terkandung dalam teks (Liu, 2012). Analisis sentimen mengkaji perasaan dan opini masyarakat terhadap suatu topik tertentu untuk mendapatkan informasi yang dapat digunakan dalam memahami, menjelaskan, dan memprediksi fenomena sosial (Pozzi et al., 2017). Opini masyarakat terhadap suatu topik dikelompokkan menjadi beberapa kelas yang mencerminkan jenis sentimen tertentu. Pada praktiknya, analisis sentimen dapat digunakan perusahaan untuk mengetahui tingkat penerimaan produk di pasar dan menciptakan strategi guna meningkatkan kualitasnya. Selain itu, analisis sentimen dapat membantu pembuat kebijakan seperti pemerintah maupun politisi dalam menganalisa persepsi masyarakat terhadap suatu kebijakan, layanan publik, dan berbagai isu politik (Prabowo & Thelwall, 2009).

2.2.2 Text Mining

Text mining merupakan teknik pengolahan data teks untuk menemukan informasi penting dari kumpulan teks yang belum terstruktur. *Text mining* menggabungkan keilmuan *data mining*, *natural language processing*, *machine learning*, dan *information retrieval* (Feldman & Sanger, 2007). Proses pencarian informasi di dalamnya melibatkan tahapan *text preprocessing*, representasi teks, dan ekstraksi teks sebelum data dianalisis lebih lanjut (Aggarwal & Zhai, 2012).

2.2.3 Text Preprocessing

Text preprocessing merupakan proses pengolahan data berupa teks sebelum dianalisis yang bertujuan untuk meningkatkan keakuratan hasil (Hickman et al., 2022). *Preprocessing* digunakan untuk mempersiapkan data agar dapat digunakan pada proses analisis dan pengolahan lanjutan secara efektif, seperti pengolahan bahasa alami yang meliputi proses membersihkan dan menata data agar lebih sesuai dan efisien untuk dianalisis (Rianto et al., 2021). Berikut merupakan tahapan *text preprocessing* yang dilakukan.

1. *Data Cleaning*

   Data cleaning adalah proses pengolahan data kotor menjadi data bersih yang tepat dan valid sehingga dapat meningkatkan kualitas data untuk disimpan atau dianalisis lebih lanjut. Prosesnya meliputi perbaikan data tidak akurat, tidak lengkap, dan tidak konsisten (Baker, 2022).

2. *Case Folding*

   Case folding merupakan tahapan normalisasi teks dalam Unicode (standar internasional representasi teks) yang memetakan semua karakter ke huruf besar (*uppercase*) lalu ke huruf kecil (*lowercase*) sehingga setiap kalimat menjadi seragam tanpa kapital (Korpela, 2006).

3. Normalisasi Slang

   Normalisasi slang merupakan proses mengubah kata slang yang berisi kosakata informal, tidak baku, dan sering digunakan dalam percakapan sehari-hari menjadi kata baku untuk menghilangkan variasi pada representasi teks sehingga dapat lebih dipahami oleh algoritma analisis sentimen (Yessy Asri & Dr. Dra. Dwina Kuswardani, 2024).

4. Tokenization

   Tokenization merupakan proses dasar dimana teks dipecah menjadi token. Token dapat berupa kata, subkata, karakter, atau bahkan frasa sesuai kebutuhan dan jenis tugasnya (Barua et al., 2024).

5. *Stopword Removal*

   Stopword removal merupakan tahapan menghilangkan kata bersifat umum dengan makna yang tidak penting dalam suatu teks untuk mengurangi *noise* dan mengurangi kompleksitas komputasi dalam *natural language processing* guna meningkatkan hasil analisis yang lebih efektif dan efisien (Moscato et al., 2025)

6. *Stemming*

   Stemming merupakan teknik penting dalam *natural language processing* yang berfungsi membuat kata berimbuhan menjadi bentuk dasarnya. *Stemming* mengubah kata menjadi bentuk dasarnya sesuai konteks dan kelas kata sehingga lebih akurat maknanya (LLC, 2025).

2.2.4 Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency (TF) berfungsi untuk mengetahui frekuensi kata tertentu muncul dalam sebuah dokumen dari keseluruhan total kata dalam dokumen tersebut (Padmaja et al., 2024). Sementara itu, *Inverse Document Frequency* (IDF) berfungsi mengetahui frekuensi suatu kata yang jarang muncul di seluruh dokumen dalam sistem untuk menentukan tingkat kepentingan relatif kata tersebut. Setiap bobot kata dievaluasi dengan teknik TF-IDF untuk memutuskan apakah kata tersebut sering muncul dan cukup representatif atau jarang muncul tetapi memiliki makna penting. Hasil dari pembobotan dipresentasikan dalam bentuk vektor fitur TF-IDF yang

mengubah setiap dokumen menjadi kumpulan nilai numerik berdasarkan kata-kata yang terkandung di dalamnya. Melalui penggabungan TF-IDF, penilaian terhadap relevansi suatu kata lebih akurat. Dalam praktiknya, nilai TF-IDF setiap kata direpresentasikan dalam bentuk kolom atau fitur. Hal tersebut membantu penyaringan kata-kata dan mengurangi dimensi data sehingga proses analisis lebih efisien (Padmaja et al., 2024).

2.2.5 Supervised Learning

Supervised learning adalah pendekatan dalam *machine learning* yang memanfaatkan data berlabel untuk melatih algoritma agar menghasilkan model terbaik. Model tersebut digunakan untuk memprediksi *output* dari data baru (Wang, 2024).

2.2.6 Unsupervised Learning

Unsupervised learning adalah salah satu pendekatan pada *machine learning* yang menggunakan algoritma untuk menemukan pola tersembunyi dalam kumpulan data tanpa label atau kategori. Pendekatan ini mengelompokkan data berdasarkan kesamaan dan perbedaannya (Geetha & Sendhilkumar, 2023). Pendekatan ini juga dianggap lebih kuat karena mampu mengenali pola tanpa bergantung pada kategori tertentu dan berpotensi membuat hasil yang lebih akurat karena tidak bergantung pada asumsi bahwa pelabelan data sudah benar (Phasinam et al., 2022). Salah satu algoritma *unsupervised learning* yang banyak digunakan adalah *K-Means Clustering*.

2.2.7 *K-Means Cluster*ing

K-Means Clustering merupakan metode pengelompokkan data yang membagi data menjadi sejumlah *cluster* (k) yang telah ditentukan sebelumnya (Prianto & Bunyamin, 2020). *K-Means* memiliki keunggulan berupa kesederhanaan, kemudahan implementasi, dan efisiensi dalam mengelompokkan kumpulan data berukuran besar. Akan tetapi, metode ini memiliki kelemahan seperti sensitif terhadap penentuan jumlah *cluster* sehingga pada penelitian ini digunakan metode penentuan jumlah *cluster*, yaitu metode *Elbow*.

2.2.8 Tahapan *K-Means Cluster*ing

Pada metode *K-Means Clustering* terdapat beberapa tahapan pelaksanaan. Menurut (Prianto & Bunyamin, 2020), berikut merupakan tahapan metode *K-Means Clustering*.

1. Menentukan jumlah *cluster* (k)

   Jumlah *cluster* (k) atau banyaknya kelompok yang akan dibentuk ditentukan terlebih dahulu. Penentuan nilai k mempertimbangkan aspek teoritis maupun konseptual sesuai kebutuhan penelitian.
2. Menentukan *centroid* awal

   Centroid awal atau titik pusat *cluster* ditentukan dengan random dari data yang tersedia sebanyak jumlah *cluster* (k) yang telah ditentukan. *Centroid* ini akan menjadi titik awal dalam proses pengelompokan data.
3. Menghitung jarak setiap data ke *centroid*

   Setiap data dihitung jaraknya terhadap setiap *centroid*. Perhitungan jarak menggunakan *Euclidean Distance* untuk mengetahui kedekatan antara data dengan *centroid* (Ragavan, 2025).
4. Mengelompokkan data ke dalam *cluster* terdekat

   Setiap data dialokasikan pada *cluster* yang mempunyai jarak paling dekat dengan *centroid*. Proses ini menggunakan pendekatan *hard K-Means* dimana satu data dinyatakan menjadi anggota *cluster* berdasarkan kedekatan sifatnya terhadap *centroid*.
5. Memperbarui *centroid* dan iterasi

   Setelah semua data dikelompokkan, posisi *centroid* diperbarui dengan menghitung rata-rata dari seluruh data dalam masing-masing *cluster*. Kemudian, langkah perhitungan jarak dan pembaruan *centroid* diulangi kembali sampai posisi *centroid* tidak berubah (Tan et al., 2014).

2.2.9 Penentuan Jumlah *Cluster* dengan Metode Elbow

Jumlah *cluster* optimal ditentukan melalui titik *Elbow* yang merupakan tahapan penting dalam metode *K-Means*. Metode *Elbow* adalah salah satu pendekatan yang digunakan untuk menentukan jumlah *cluster* yang sesuai dengan dataset tertentu. Dalam menentukan jumlah *cluster* yang optimal, metode *Elbow* menggunakan nilai *Within-Cluster Sum of Squares* (WCSS), yaitu ukuran kedekatan masing-masing data dalam satu *cluster* (Noring et al., 2024).

2.2.10 Aplikasi Instagram

Instagram adalah media sosial yang menampilkan visual dengan berbagai data melalui konten gambar atau video, *caption*, *hashtag*, dan komentar. Hal ini membuat instagram sebagai sumber data yang dapat digunakan dalam penelitian seperti analisis sentimen. Peneliti dapat memahami

opini dan konteks interaksi pengguna secara menyeluruh. Namun, akses terhadap data instagram dalam beberapa tahun ini menjadi semakin terbatas karena kebijakan *platform* sehingga menjadi tantangan dalam mengumpulkan data. Pengumpulan dan analisis data dapat dilakukan melalui teknik *machine learning* maupun secara manual. Analisis tersebut dapat menggunakan metode kuantitatif maupun kualitatif. (Quan-Haase & Sloan, 2022). Dengan demikian, pemilihan metode yang tepat serta pemahaman karakteristik *platform* menjadi faktor penting dalam menghasilkan analisis sentimen yang akurat dan relevan.

2.2.11 Aplikasi YouTube

YouTube merupakan *platform* video yang menggabungkan elemen teks, audio, dan visual dalam satu ruang digital. Pengguna dapat berbagi tontonan dan berinteraksi melalui komentar, *like*, *dislike*, dan *share*. Lebih dari dua miliar pengguna bulanan di lebih dari 100 negara dan 80 konten bahasa membuat *platform* ini menghasilkan banyak data dari video teks. Hal ini membuat YouTube menjadi sumber data berharga untuk penelitian karena dapat mengamati perilaku opini dan interaksi pengguna tanpa mengganggunya. Dalam analisis media sosial, data YouTube digunakan untuk memahami sentimen, sikap, dan opini melalui teknik seperti analisis sentimen pada komentar publik, pemeriksaan jumlah tayangan video, serta pendekatan canggih dengan *machine learning* untuk memudahkan pengumpulan dan pengolahan data. YouTube sendiri memiliki tantangan seperti isu privasi, etika, kompleksitas data, dan potensi misinformasi. Akan tetapi, YouTube tetap memiliki signifikansi tinggi dalam kemampuannya membentuk opini publik, menyampaikan informasi, serta menyediakan sampel data yang mencerminkan preferensi dan interaksi masyarakat secara global (Quan-Haase & Sloan, 2022).

BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Objek pada penelitian ini adalah opini masyarakat terhadap kebijakan kewajiban sertifikasi halal di Indonesia yang diperoleh dari media sosial Instagram dan YouTube. Penelitian ini menggunakan data komentar yang berisi tanggapan masyarakat terhadap kebijakan tersebut. Fokus penelitian ini adalah mengidentifikasi dan mengelompokkan opini masyarakat ke dalam kategori sentimen tertentu menggunakan metode *K-Means Clustering*.

3.2 Jenis Data

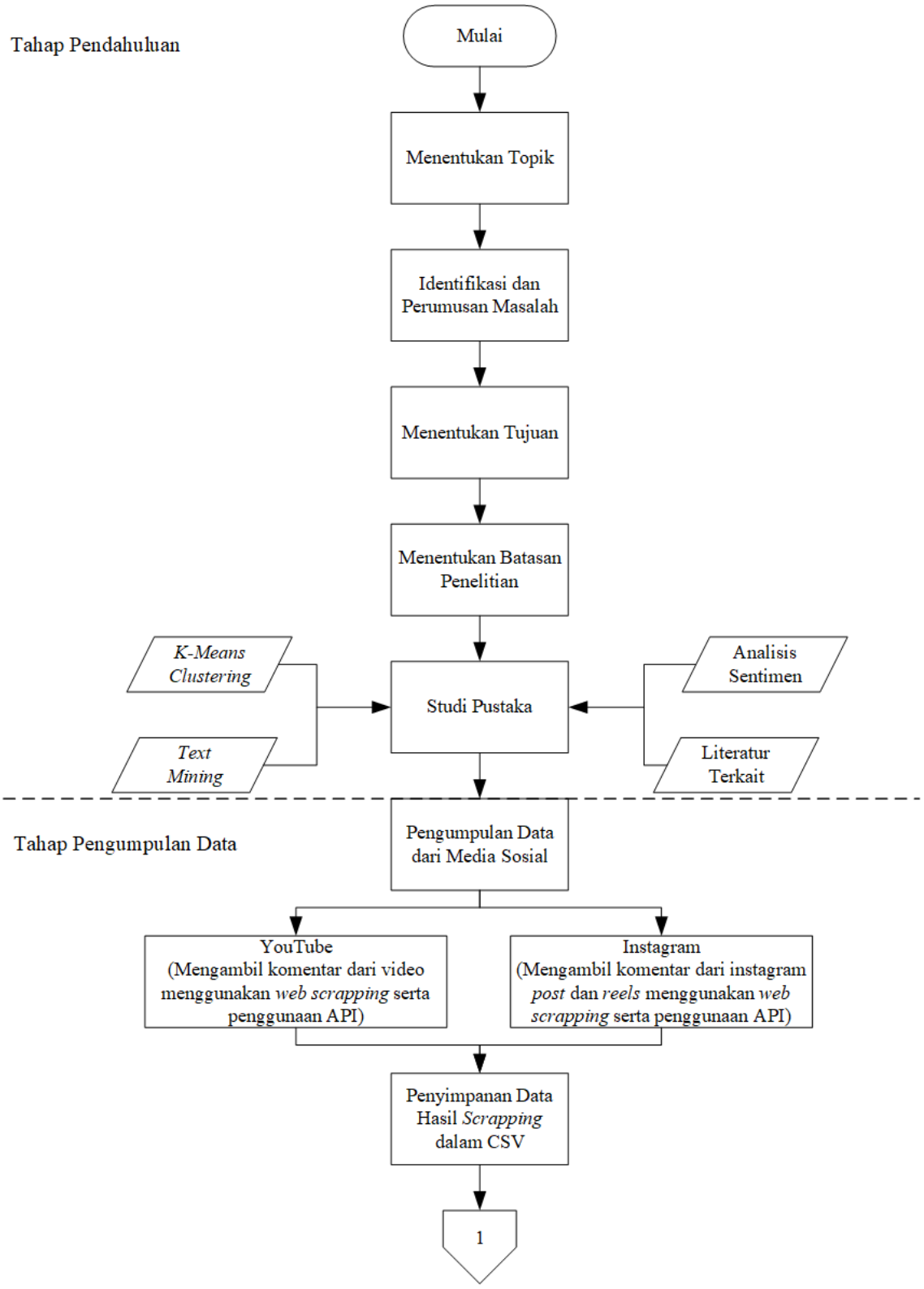
Jenis data yang digunakan adalah data sekunder. Data berupa komentar dikumpulkan dari sumber yang telah tersedia, yaitu *platform* media sosial Instagram dan YouTube. Data tersebut berisikan teks opini pengguna yang menunjukkan persepsi, sikap, dan sentimen terhadap kebijakan kewajiban sertifikasi halal. Data sekunder ini dipilih karena memiliki kemampuan dalam merepresentasikan opini publik secara luas dan terkini sehingga relevan untuk analisis sentimen dengan *text mining*.

3.3 Metode Pengumpulan Data

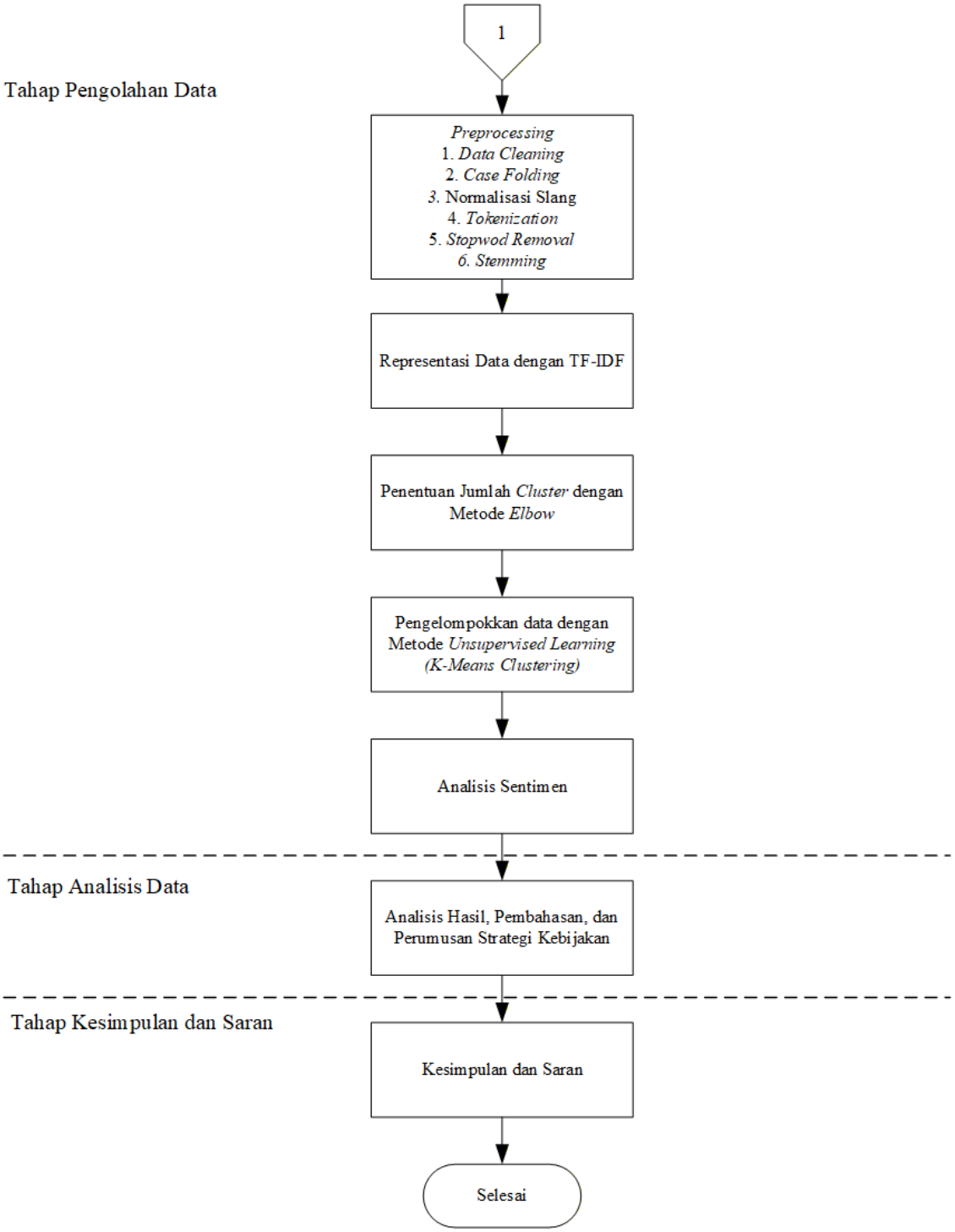
Pengumpulan data menggunakan teknik *scraping* dari *platform* Instagram dan YouTube. Proses *scraping* dilakukan menggunakan *web scraping* serta penggunaan *Application Programming Interface* (API) yang dimiliki *platform* tersebut.

3.4 Alur Penelitian

Berikut merupakan alur yang digunakan dalam penelitian ini:



Gambar 3. 1 Diagram Alur Penelitian Bagian 1



Gambar 3. 2 Diagram Alur Penelitian Bagian 2

Berdasarkan diagram alur penelitian di atas, berikut merupakan penjelasan:

1. Tahap Pendahuluan

Tahap ini menjadi dasar awal penelitian yang memuat penyusunan rancangan penelitian dan tujuan penelitian.

a. Mulai

   Menunjukkan permulaan penelitian yang menandai dimulainya penelitian.

b. Menentukan Topik

   Tahap penelitian yang dilakukan pertama kali. Topik yang diangkat adalah analisis sentimen pada kebijakan kewajiban sertifikasi halal di Indonesia.

c. Identifikasi dan Perumusan Masalah

   Tahapan ini berfungsi untuk mengenali permasalahan dengan melihat kondisi dan situasi yang secara tidak langsung. Berdasarkan hasil identifikasi, diketahui bahwa terdapat kebijakan baru pada kewajiban sertifikasi halal untuk para pelaku usaha di Indonesia. Oleh karena itu, perlu dilakukan analisis sentimen untuk mengetahui tanggapan masyarakat dan pemahaman masyarakat mengenai kebijakan tersebut. Selanjutnya, penulis membuat rumusan masalah yang akan menjadi dasar terlaksananya penelitian.

d. Menentukan Tujuan

   Pada tahapan ini, dijelaskan beberapa hal dalam penelitian yang ingin dicapai. Tujuan penelitian disusun untuk menjawab pertanyaan penelitian pada rumusan masalah.

e. Menentukan Batasan Penelitian

   Batasan penelitian dibuat untuk menjaga agar fokus penelitian tetap terarah dan pembahasan tidak meluas ke hal-hal di luar kebutuhan penelitian. Tahap ini menjelaskan ruang lingkup yang menjadi batasan penelitian.

f. Studi Pustaka

   Studi pustaka digunakan untuk meninjau teori, konsep, manfaat yang berkaitan dengan penelitian. Selain itu, terdapat sejumlah penelitian terdahulu yang digunakan sebagai referensi dalam mengidentifikasi celah penelitian dan menentukan metode yang sesuai.

2. Tahap Pengumpulan Data

Tahapan ini berisi proses pengumpulan data mentah yang sesuai dengan topik penelitian dari beberapa *platform* media sosial.

a. *Scraping* Data dari Instagram dan YouTube

   Data komentar dikumpulkan secara otomatis dari *platform* media sosial Instagram dan YouTube. Pengumpulan data menggunakan teknik *web scraping* serta *Application Programming Interface* (API) yang disediakan *platform* tersebut.

b. Penyimpanan Data

Selanjutnya, data yang telah berhasil terkumpul disimpan dalam bentuk CSV untuk digunakan pada tahap pengolahan data.

3. Tahap Pengolahan Data

Tahap ini meliputi rangkaian proses penting untuk mengolah data agar dapat dianalisis pada tahap berikutnya.

a. *Preprocessing*

Tahap awal pada pengolahan data yang berfungsi untuk membersihkan dan menyiapkan data sebelum dianalisis lebih lanjut. Proses ini terdiri dari beberapa langkah, yaitu *data cleaning* untuk menghapus karakter yang tidak diperlukan dalam teks, *case folding* untuk mengubah teks menjadi huruf kecil agar seragam, normalisasi slang untuk mengubah kata slang menjadi kata baku, *tokenization* untuk memecah kalimat menjadi token yang berupa frasa/kata-kata, *stopword removal* untuk menghapus kata-kata umum yang tidak memiliki makna signifikan, dan *stemming* untuk mengubah setiap kata menjadi kata dasar tanpa imbuhan.

b. Representasi Data dengan TF-IDF

Selanjutnya, data bersih diubah menjadi representasi numerik menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF).

c. Penentuan Jumlah *Cluster* dengan Metode *Elbow*

Pada tahap ini dilakukan penentuan jumlah *cluster* yang optimal menggunakan metode *Elbow*. Metode ini melihat titik siku pada grafik untuk menentukan jumlah *cluster* terbaik.

d. Pengelompokkan Data dengan *K-Means Clustering*

Data akan dikelompokkan ke dalam setiap *cluster* yang ada berdasarkan kemiripan atau karakteristik teks menggunakan algoritma *K-Means*.

e. Analisis Sentimen

Tahap akhir dari pengolahan data ini digunakan untuk mengidentifikasi sentimen dari setiap data yang telah dikelompokkan. Hal ini bertujuan untuk mengetahui kecenderungan opini masyarakat Indonesia terhadap topik yang diteliti.

4. Tahap Analisis Data

Seluruh hasil yang diperoleh dari proses pengolahan data selanjutnya akan diinterpretasikan dan dianalisis secara mendalam. Hasil analisis data tersebut akan dibahas dengan mengacu pada tujuan penelitian dan landasan teori yang digunakan sehingga dapat mengidentifikasi

pola yang muncul dan menjawab rumusan masalah serta memberikan rekomendasi perumusan strategi kebijakan kewajiban sertifikasi halal bagi pihak terkait.

5. Tahap Kesimpulan dan Saran

   Setelah memperoleh hasil analisis dan pembahasan, kesimpulan akan dirumuskan untuk menjawab pertanyaan penelitian yang terdapat dalam rumusan masalah. Kemudian, saran yang relevan akan diberikan sebagai rekomendasi bagi pihak terkait dan penelitian selanjutnya.

3.5 Perhitungan dalam Pembobotan TF-IDF

Nilai TF-IDF didapatkan dari hasil perkalian antara frekuensi kemunculan kata (TF) dengan tingkat kepentingannya dalam dokumen (IDF), berikut merupakan formula perhitungannya (Padmaja et al., 2024):

$$Term\ frequency\ (t) = \frac{Jumlah\ kemunculan\ t\ dalam\ dokumen}{Jumlah\ total\ kata\ dalam\ dokumen} \quad (3.1)$$

$$Inverse\ Document\ Frequency\ (t) = \frac{Jumlah\ total\ dokumen}{Jumlah\ dokumen\ yang\ mengandung\ t} \quad (3.2)$$

3.6 Perhitungan dalam Penentuan Jumlah *Cluster* dengan Elbow

Penentuan jumlah *cluster* menggunakan metode *Elbow* dilakukan dengan membandingkan nilai *Within-Cluster Sum of Squares* (WCSS) pada beberapa nilai k. Nilai k akan dipilih pada titik saat penurunan WCSS mulai melambat atau membentuk titik siku. Berikut formula perhitungan WCSS (Rafika et al., 2025):

$$WCSS\ = \sum_{i=1}^{k} \sum x\ \in C_i\ \|x - \mu_i\|^2 \quad (3.3)$$

Keterangan:
k = jumlah *cluster*
C_i = *cluster* ke-i
x = data pada *cluster* C_i
μ_i = *centroid* atau pusat *cluster* ke-i
$\|x - \mu_i\|^2$ = jarak antara data x dengan *centroid*

3.7 Perhitungan *K-Means Cluster*ing

Pada pengelompokkan data menggunakan metode *K-Means Clustering*, terdapat dua formula yang digunakan, yaitu perhitungan jarak menggunakan *Euclidean Distance* dan perhitungan pembaruan *centroid*. Berikut merupakan persamaan *Euclidean Distance* yang digunakan untuk data berdimensi banyak atau n-dimensi (Ragavan, 2025):

$$d = \sqrt{\sum_{i=1}^{n}(x_{2i} - x_{1i})^2} \quad (3.4)$$

Keterangan:
d = jarak antara dua data
x_{1i} = nilai fitur ke-i pada data pertama
x_{2i} = nilai fitur ke-i pada data kedua (*centroid*)
i = indeks fitur atau dimensi data
n = jumlah seluruh fitur atau dimensi data

Kemudian, berikut merupakan formula perhitungan untuk pembaruan *centroid* (Tan et al., 2014):

$$c_i = \frac{\sum x_j \epsilon {C_i}^{x_j}}{|C_i|} \quad (3.5)$$

Keterangan:
c_i = *centroid* atau titik pusat *cluster* ke-i
C_i = jumlah data yang termasuk ke dalam *cluster* ke-i
x_j = data ke-j yang menjadi anggota *cluster*
i = indeks *cluster*
j = indeks data dalam *cluster*

BAB IV
PENGUMPULAN DAN PENGOLAHAN DATA

4.1 Pengumpulan Data

4.1.1 Aplikasi YouTube dan Instagram

Pengumpulan data dalam penelitian ini dilakukan secara otomatis menggunakan teknik *web scraping* pada *platform* YouTube dan Instagram. Fokus pengumpulan data membahas mengenai kewajiban sertifikasi halal di Indonesia. Periode data untuk kedua sumber adalah dari April 2024 sampai April 2026. Berikut merupakan tahapan pengumpulan data:

1. Penentuan Kata Kunci dan Sumber Data

   Tahap awal pengumpulan data dilakukan dengan menentukan kata kunci sesuai topik penelitian untuk menemukan dan mengidentifikasi video dan unggahan platform YouTube dan Instagram agar konten dan komentar yang akan dikumpulkan relevan dengan topik penelitian. Kata kunci yang digunakan seperti ”sertifikasi halal”, ”label halal”, ”BPJPH”, ”SEHATI”, ”UMK halal”, dan ”sertifikat halal gratis”.

2. Rentang Waktu Pengumpulan Data

   Data komentar yang digunakan dalam penelitian ini dikumpulkan dalam rentang waktu April 2024 sampai April 2026. Rentang waktu tersebut digunakan karena berdekatan dengan periode pemberlakuan kebijakan wajib sertifikasi halal sehingga diskusi publik mengenai topik tersebut sedang meningkat, terutama untuk pelaku usaha kecil. Dengan demikian, data yang didapatkan diharapkan dapat memberi gambaran terhadap respon masyarakat dengan lebih aktual.

3. Pengumpulan Data YouTube

   Pengumpulan data dari YouTube dilakukan dengan mengambil komentar pada beberapa video yang membahas sertifikasi halal di Indonesia. Data yang dikumpulkan berisi kolom tanggal komentar, nama pengguna, komentar, likes, jenis komentar, dan tautan sumber video. Komentar yang telah berhasil di-*scraping* lalu disimpan dalam bentuk CSV.

4. Pengumpulan Data Instagram

   Pengumpulan data dari Instagram dilakukan dengan mengambil komentar pada beberapa unggahan dan reels yang membahas sertifikasi halal di Indonesia. Data yang dikumpulkan memiliki atribut tanggal komentar, nama pengguna, komentar, likes, jenis komentar, dan tautan sumber unggahan. Data komentar kemudian disimpan dalam bentuk CSV untuk

diolah lebih lanjut. Pengumpulan data Instagram dilakukan untuk melengkapi data YouTube sehingga penelitian dapat menggambarkan opini secara lebih luas dari dua *platform* berbeda.

5. Penyimpanan dan Penggabungan Data

   Pada tahapan ini, data yang diperoleh dari YouTube dan Instagram digabungkan dengan struktur kolom yang terdiri dari tanggal, nama, komentar, likes, tipe, link sumber, dan asal *platform*. Berikut merupakan sampel data dari hasil penggabungan:

Tabel 4. 1 Sampel Data *Scraping*

index	tanggal	nama	komentar	likes	tipe	link_sumber	platform
1	2026-04-08 03:25:38.426 389	@MunifaEca	Sdh jalan supaja dpt duit	0	comment	https://youtu.be/yaBnRLC6AhQ?si=xQ41Kr8NgUO_qRSV	YouTube
2	2026-03-30 03:25:38.432 219	@memedcreative	Sekarang dah 2026 mana hoax	0	comment	https://youtu.be/yaBnRLC6AhQ?si=xQ41Kr8NgUO_qRSV	YouTube
3	2025-12-17T23:51:24.000Z	@pamujisugeng1	Keren nih infonya bermanfaat banget	0	comment	https://www.instagram.com/p/DSSPXzlk6h1/?igsh=MXdsMHk4eG1udWRybw==	Instagram
4	2025-12-27T17:57:42.000Z	@mohamadfitri091976	Keren halal indonesia	0	comment	https://www.instagram.com/p/DSSPXzlk6h1/?igsh=MXdsMHk4eG1udWRybw==	Instagram
5	2025-12-16T12:44:08.000Z	@nurfadillahkahf	Semoga disegerakan pelstihan penyelia halal	0	comment	https://www.instagram.com/p/DSSPXzlk6h1/?igsh=MXdsMHk4eG1udWRybw==	Instagram

4.2 Pengolahan Data

4.2.1 Pre-Processing Data

Pre-processing data teks adalah tahapan yang sangat penting dalam pengolahan data untuk membersihkan data, normalisasi data, dan membuat data mentah menjadi format yang tepat untuk dianalisis lebih lanjut menggunakan *clustering*. Data yang diperoleh dari dua sumber berbeda yaitu X dan YouTube serta telah digabung pada tahap sebelumnya akan memiliki tahapan yang sama untuk menjaga konsistensi dan kualitas data. Kode *pada pre-processing* akan ditampilkan dengan rinci pada Lampiran. Berikut beberapa tahapan yang dilakukan pada *pre-processing*.

1. *Data Cleaning*

   Tahapan ini dilakukan untuk membersihkan teks dari unsur-unsur yang tidak diperlukan dalam analisis sentimen dan pengelompokkan *cluster*. Beberapa bagian yang dihapus seperti URL atau tautan web, *mention username*, simbol hashtag, emoji, angka, tanda baca, dan spasi berlebih yang tidak berkontribusi signifikan pada makna teks.

Tabel 4. 2 *Data Cleaning*

Data Sebelum	**Setelah *Cleaning***
Apakah pemerintah Ga bisa nyari penghasilan sendiri? Selain majakin orang?	Apakah pemerintah Ga bisa nyari penghasilan sendiri Selain majakin orang

2. *Case Folding*

   Tahapan ini dilakukan untuk menyeragamkan representasi data teks. Setiap huruf kapital akan diubah menjadi huruf kecil (*lowercase*).

Tabel 4. 3 *Case Folding*

Data Sebelum	**Setelah *Case Folding***
Apakah pemerintah Ga bisa nyari penghasilan sendiri Selain majakin orang	apakah pemerintah ga bisa nyari penghasilan sendiri selain majakin orang

3. Normalisasi Slang

Tahapan ini digunakan untuk mengubah bentuk kata tidak baku, bahasa gaul, singkatan kata menjadi kata baku. Hal ini dilakukan agar makna teks lebih jelas dan seragam.

Tabel 4. 4 Normalisasi Slang

Data Sebelum	**Setelah Normalisasi Slang**
apakah pemerintah ga bisa nyari penghasilan sendiri selain majakin orang	apakah pemerintah tidak bisa mencari penghasilan sendiri selain memajaki orang

4. Tokenisasi

Tokenisasi adalah proses memecah teks menjadi bagian-bagian kecil yang disebut token. Pada tahap ini, setiap kata dalam teks akan dipisahkan dan dianggap sebagai satu token.

Tabel 4. 5 Tokenisasi

Data Sebelum	**Setelah Tokenisasi**
apakah pemerintah tidak bisa mencari penghasilan sendiri selain memajaki orang	['apakah', 'pemerintah', 'tidak', 'bisa', 'mencari', 'penghasilan', 'sendiri', 'selain', 'memajaki', 'orang']

5. Penghapusan *Stopword*

Tahapan ini digunakan untuk menghapus kata-kata umum yang kurang memiliki makna penting terhadap hasil analisis data. Proses ini dilakukan agar data menjadi lebih sederhana dan fokus pada kata-kata yang memiliki makna penting. Daftar *stopword* yang disesuaikan dalam bahasa Indonesia.

Tabel 4. 6 Penghapusan *Stopword*

Data Sebelum	**Setelah Penghapusan *Stopword***
['apakah', 'pemerintah', 'tidak', 'bisa', 'mencari', 'penghasilan', 'sendiri', 'selain', 'memajaki', 'orang']	['pemerintah', 'tidak', 'mencari', 'penghasilan', 'sendiri', 'memajaki', 'orang']

6. Proses *Stemming*

Stemming adalah proses mengubah kata berimbuhan menjadi bentuk dasarnya. Tahap ini dilakukan untuk membuat kata-kata yang memiliki makna dasar sama menjadi dapat diseragamkan. Contohnya seperti jualan menjadi jual. Dengan demikian, hasil analisis akan lebih konsisten.

Tabel 4. 7 Proses *Stemming*

Data Sebelum	**Setelah *Stemming***
['pemerintah', 'tidak', 'mencari', 'penghasilan', 'sendiri', 'memajaki', 'orang']	['pemerintah', 'tidak', 'cari', 'hasil', 'sendiri', 'pajak', 'orang']

4.2.2 Representasi Teks Menggunakan TF-IDF

Tahapan ini dilakukan untuk mengubah data komentar yang telah melalui *pre-processing* menjadi bentuk numerik untuk dapat diproses lebih lanjut pada tahap *clustering*. Proses TF-IDF menggunakan fungsi *library scikit-learn* yaitu TfidfVectorizer dengan jumlah fitur maksimal sebanyak 5000 kata, kata yang muncul kurang dari 2 kali diabaikan, dan kata yang terlalu sering muncul lebih dari 80% dokumen tidak digunakan. Tahapan ini menghasilkan matriks TF-IDF yang merepresentasikan bobot setiap kata dalam komentar dan digunakan sebagai input dalam proses *clustering*. Pembobotan menghasilkan matriks berukuran 4841 baris teks data dan 2803 kolom kata. Nilai pembobotan TF-IDF berada pada *range* 0 sampai 1 dimana 0 menunjukkan suatu kata tidak muncul atau tidak memiliki bobot pada baris teks tertentu sedangkan nilai lebih besar menunjukkan suatu kata memiliki tingkat kepentingan yang lebih tinggi dalam suatu baris teks.

Tabel 4. 8 Pembobotan TF-IDF

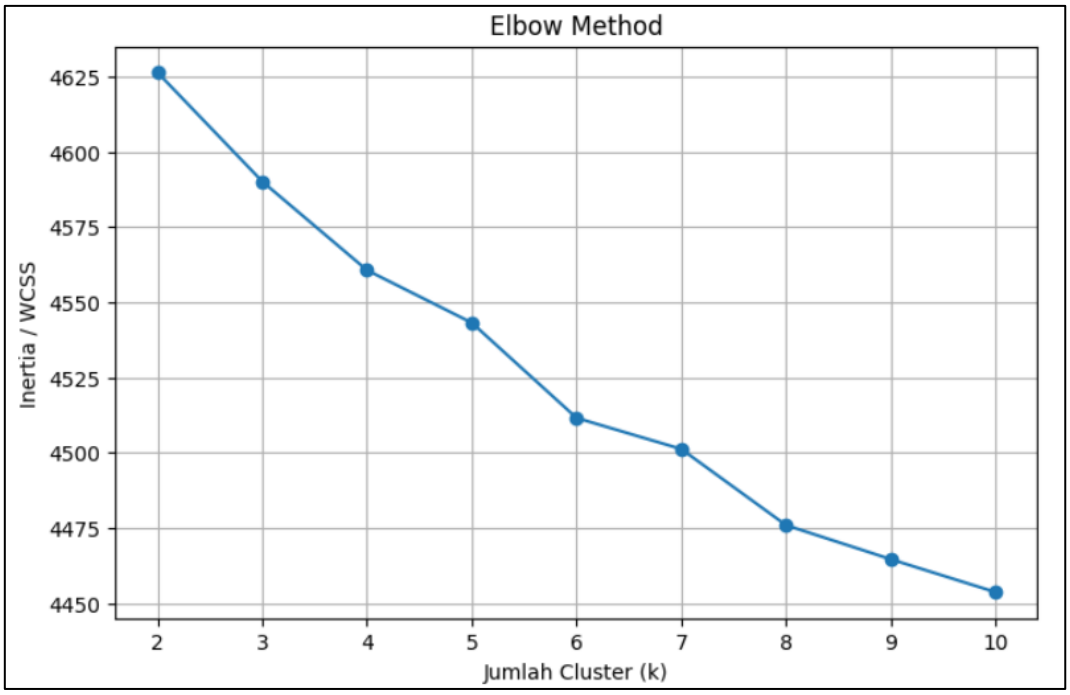
Baris	**daftar**	**hasil**	**cari**	**orang**	**pemerintah**	**sendiri**	**tidak**	**agama**	**halal**	**semua**
1	1	0	0	0	0	0	0	0	0	0
2	0	0,504	0,501	0,342	0,361	0,450	0,212	0	0	0
3	0	0	0	0,098	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0,689	0,332	0,645
5	0	0	0	0	0	0	0,154	0	0	0

4.2.3 Penentuan Jumlah *Cluster* dengan Metode Elbow

Penentuan jumlah *cluster* dilakukan menggunakan metode *Elbow* dengan melihat nilai *inertia* atau *Within Cluster Sum of Squares* (WCSS) pada beberapa jumlah *cluster* (k). Berdasarkan grafik yang diperoleh, nilai WCSS mulai menurun seiring bertambahnya jumlah *cluster* mulai dari k = 2 sampai k = 10. Penurunan yang cukup signifikan terjadi pada k = 6, selanjutnya penurunan nilai WCSS lebih kecil dan cenderung melandai. Dengan demikian, jumlah *cluster* optimal yang digunakan dalam penelitian ini adalah k = 6 karena pada titik tersebut terjadi perubahan pola penurunan signifikan serta menunjukkan bahwa penambahan *cluster* selanjutnya tidak memberikan penurunan nilai WCSS yang signifikan. Berikut merupakan perhitungan selisih nilai WCSS dan grafik metode elbow yang menjadi dasar pembentukan jumlah *cluster*.

Tabel 4. 9 Selisih/Gap Nilai WCSS Pembentukan Jumlah *Cluster*

k (Jumlah *Cluster*)	**Nilai WCSS/*Inertia***	**Selisih/Gap**	**Presentase Penurunan**
2	4626,28	-	-
3	4590,07	36,21	0,78
4	4560,60	29,47	0,64
5	4543,14	17,46	0,38
6	4511,53	31,61	0,7
7	4501,19	10,34	0,23
8	4475,91	25,28	0,56
9	4464,59	11,32	0,25
10	4453,66	10,93	0,24



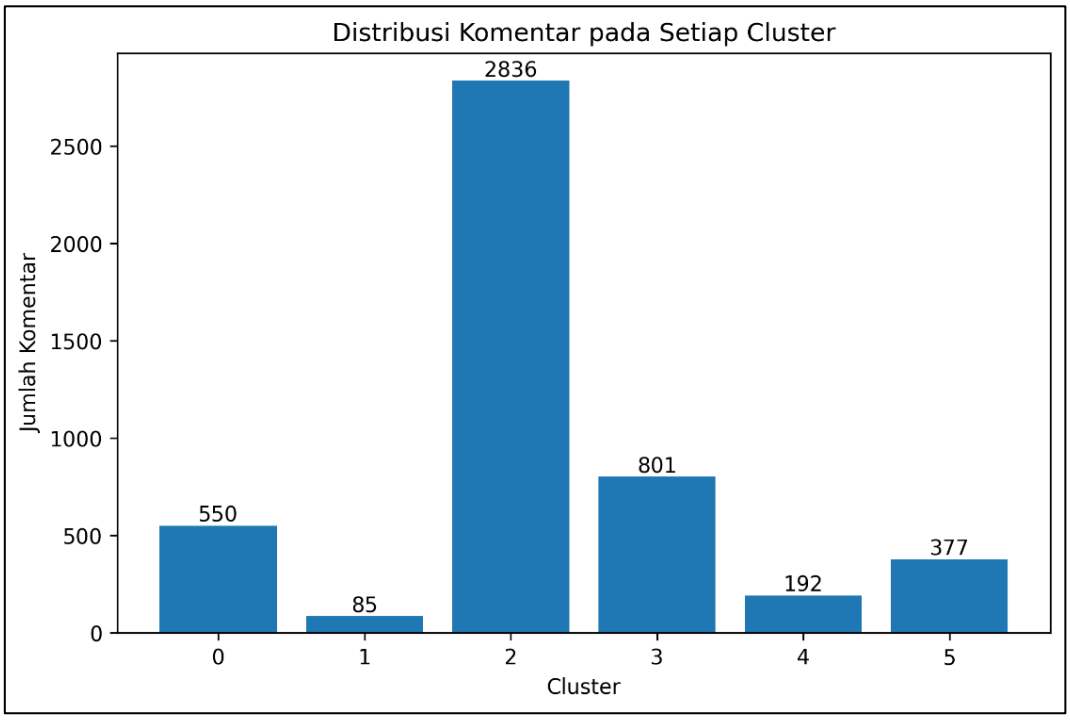
Gambar 4. 1 Grafik Metode *Elbow*

Berdasarkan perhitungan selisih/gap nilai WCSS dan grafik metode elbow, jumlah *cluster* yang memiliki penurunan signifikan adalah k=6. Meskipun pada k=2 juga terjadi penurunan besar akan tetapi kondisi tersebut mencerminkan penurunan awal karena pembagian data menjadi 2 *cluster* sehingga belum dapat memberikan pemisahan data yang optimal. Sedangkan, setelah k=6 tidak ada penurunan yang signifikan atau lebih besar lagi sehingga menandakan bahwa penambahan *cluster* tidak lagi memberikan peningkatan kualitas pengelompokkan yang besar.

4.2.4 Pengelompokkan Data dengan *K-Means Clustering*

Data komentar yang telah direpresentasikan dalam bentuk numerik menggunakan metode TF-IDF selanjutnya akan dikelompokkan dengan algoritma *K-Means Clustering*. Proses pengelompokkan dilakukan dengan jumlah *cluster* sebanyak 6 dari hasil metode *Elbow*. Proses *clustering* menggunakan parameter n_*Cluster*=6, random_state=42 untuk mengatur agar hasil *clustering* tetap sama setiap kali dilakukan *run*, dan n_init=10 untuk menunjukkan bahwa algoritma *K-Means* telah dijalankan sebanyak 10 kali dengan pembaruan titik pusat *cluster*. Kemudian, hasil terbaik dipilih berdasarkan nilai *inertia* terkecil. Ketiga parameter tersebut

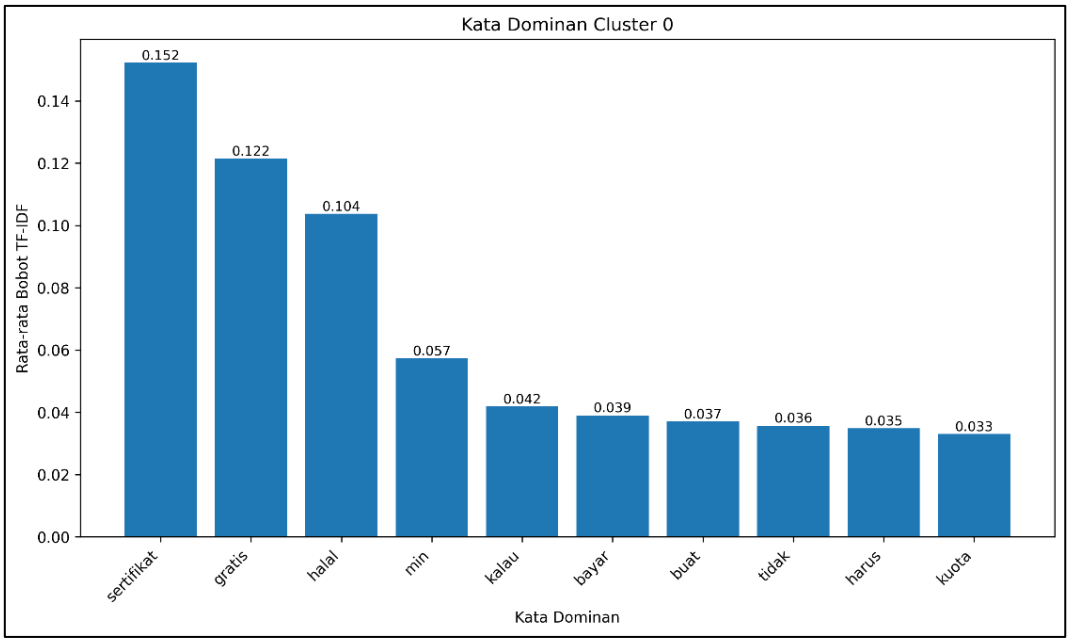
digunakan untuk memastikan agar proses *clustering* stabil dan hasil pengelompokkan tidak berubah-ubah. Berikut merupakan jumlah setiap *cluster* yang berhasil dikelompokkan:



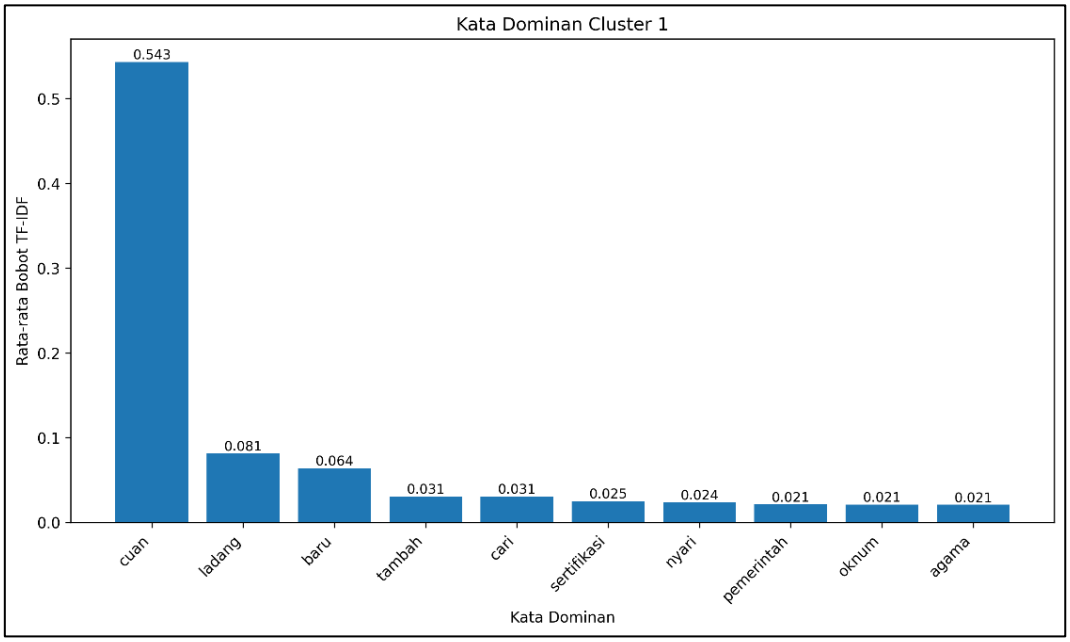
Gambar 4. 2 Distribusi Jumlah Komentar Setiap *Cluster*

4.2.5 Profilisasi dan Labeling *Cluster*

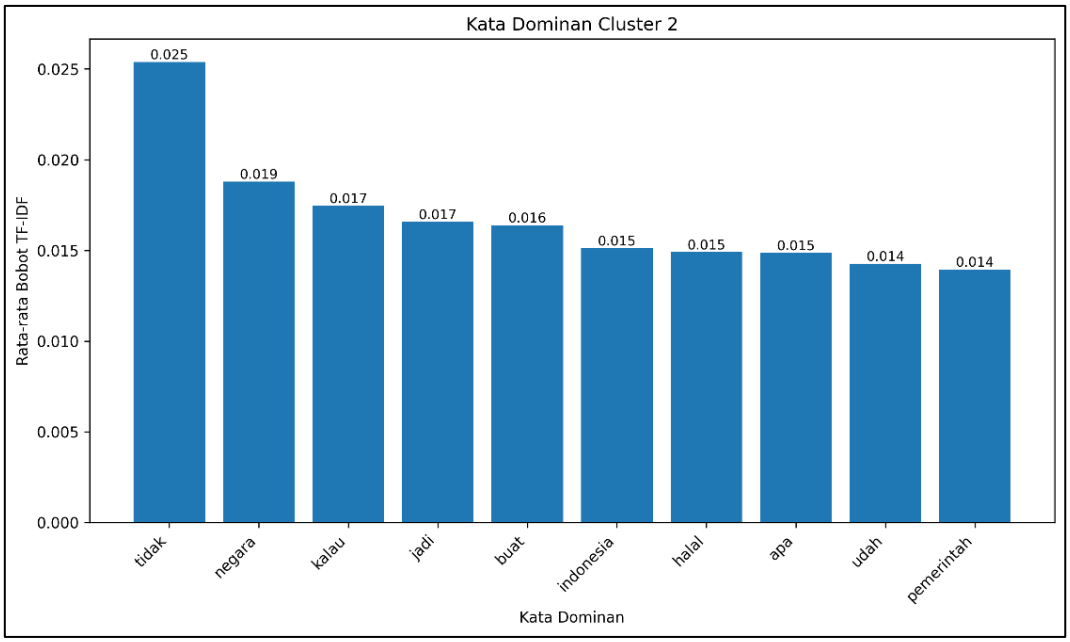
Profilisasi *cluster* dilakukan untuk memahami karakteristik dari setiap kelompok komentar yang ada sehingga dapat dilakukan pelabelan untuk masing-masing *cluster*. Tahapan ini menganalisis setiap *cluster* berdasarkan jumlah komentar dan kata dominan yang didapatkan dari rata-rata bobot TF-IDF tertinggi. Kata dominan digunakan sebagai dasar untuk menentukan labeling atau topik utama yang diperbincangkan pada setiap *cluster*. Berikut merupakan grafik dari 10 kata dominan pada masing-masing *cluster*:



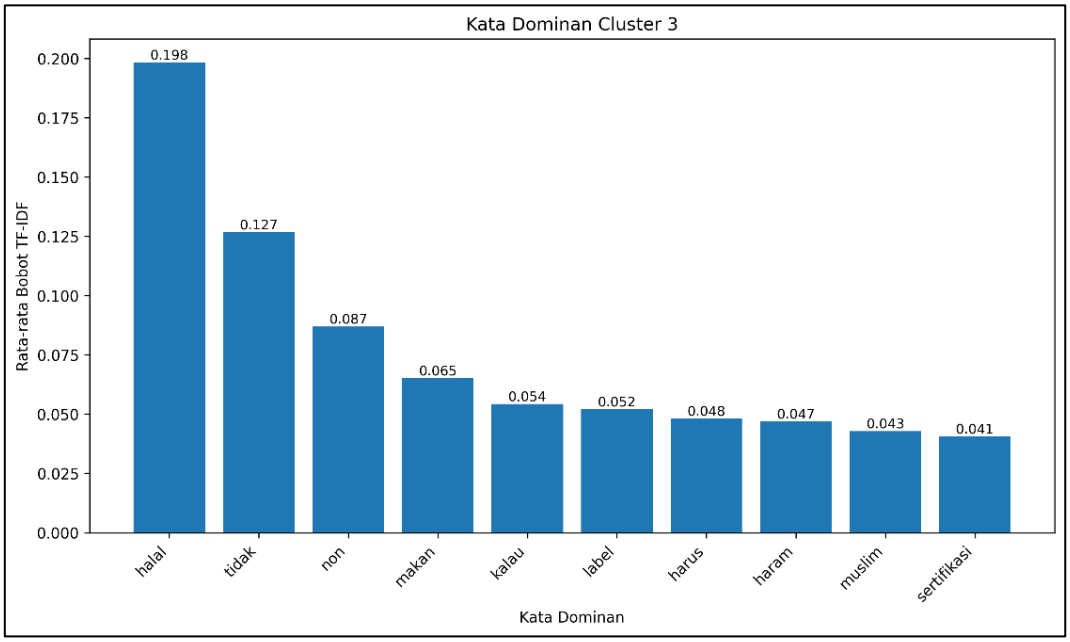
Gambar 4. 3 Grafik Kata Dominan *Cluster* 0



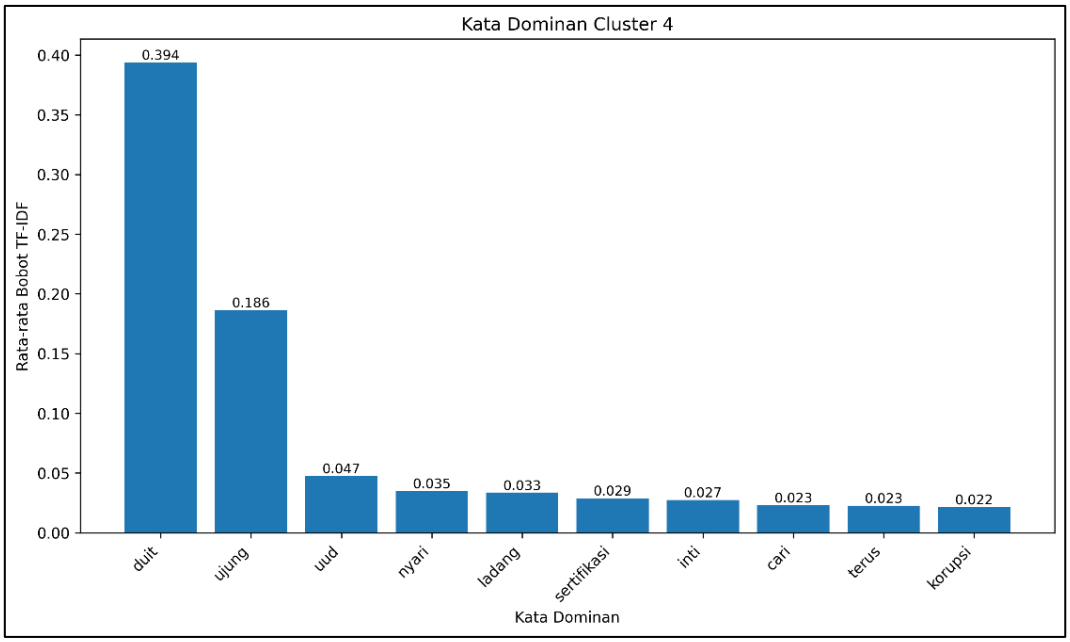
Gambar 4. 4 Grafik Kata Dominan *Cluster* 1



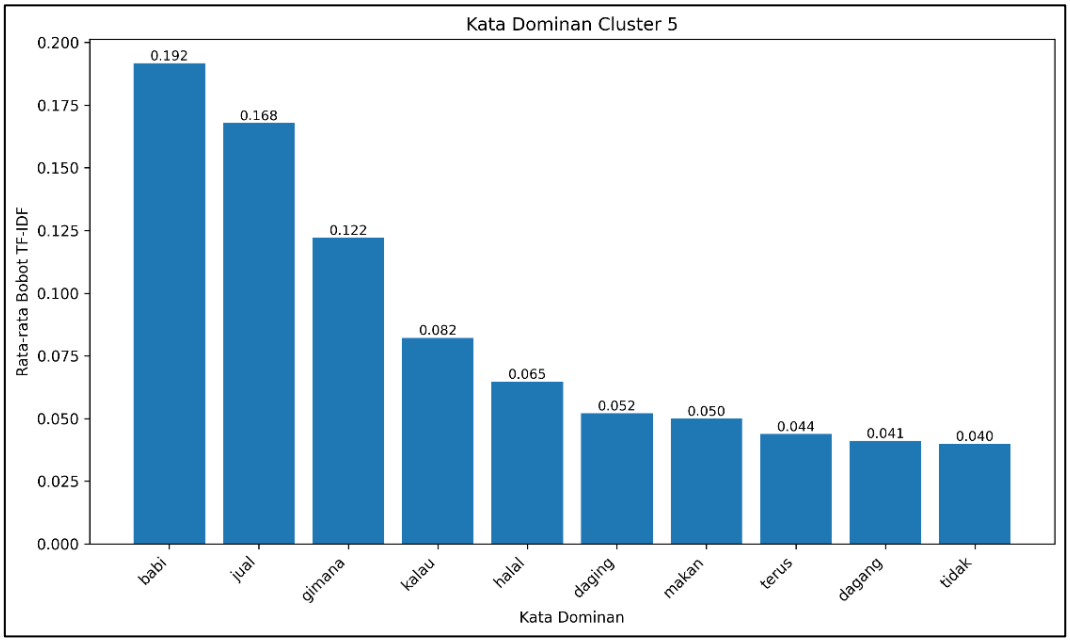
Gambar 4. 5 Grafik Kata Dominan *Cluster* 2



Gambar 4. 6 Grafik Kata Dominan *Cluster* 3



Gambar 4. 7 Grafik Kata Dominan *Cluster* 4



Gambar 4. 8 Grafik Kata Dominan *Cluster* 5

Berdasarkan visualisasi setiap grafik di atas, masing-masing *Cluster* memiliki karakteristik yang berbeda. Setiap grafik menampilkan 10 kata dominan berdasarkan rata-rata bobot TF-IDF tertinggi yang merepresentasikan pembahasan dalam masing-masing *cluster*. *Cluster* 0 didominasi kata yang membahas sertifikasi halal dan program gratis. *Cluster* 1 didominasi kata yang membahas anggapan bahwa kebijakan sertifikasi halal dapat menjadi

peluang keuntungan atau dimanfaatkan oleh oknum tertentu yang merugikan masyarakat. *Cluster* 2 berisi kata-kata yang membahas kebijakan, agama, dan masyarakat. *Cluster* 3 berisi pembahasan mengenai label halal pada produk makanan. Kemudian, *cluster* 4 menampilkan kata-kata yang berisi pembahasan mengenai dugaan korupsi dan pungutan. Sementara itu, *cluster* 5 membahas pertanyaan masyarakat terkait produk nonhalal seperti babi dan makanan haram lainnya.

Selanjutnya, kata dominan yang muncul pada masing-masing *cluster* akan dianalisis lebih lanjut sebagai dasar dalam profilisasi dan labeling *cluster*. Tahapan ini digunakan untuk memberikan nama atau label yang sesuai dengan karakteristik pembahasan setiap *cluster*. Hal ini menunjukkan bahwa hasil pengelompokkan data tidak hanya menampilkan kemiripan data tetapi juga memberikan gambaran topik utama serta kecenderungan opini masyarakat. Berikut merupakan tabel profilisasi dan labeling *cluster*:

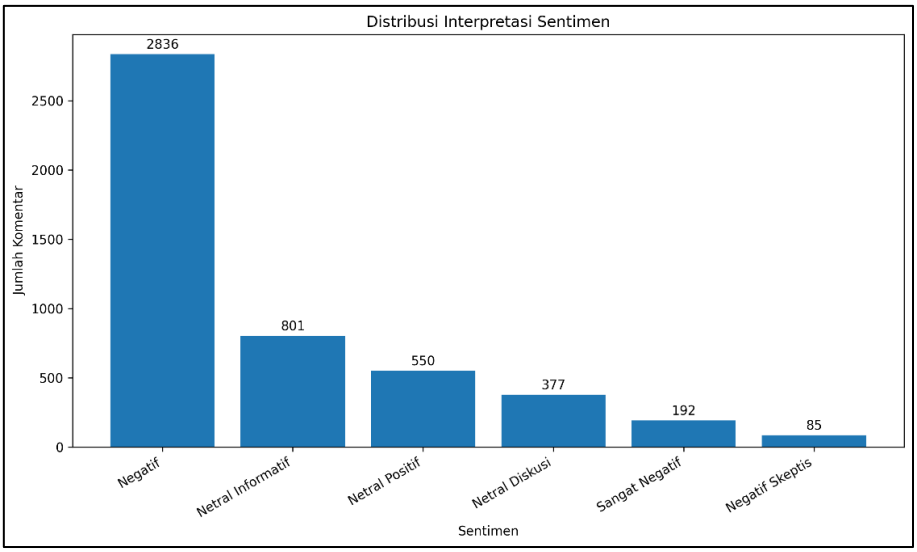
Tabel 4. 10 Profilisasi dan Labeling *Cluster*

Cluster	**Kata Dominan**	**Contoh Komentar**	**Labeling**	**Sentimen**
0	sertifikat, gratis, halal, min, kalau, bayar, buat, tidak, harus, kuota, daftar, usaha, umkm, dapat, sertifikasi	”Mohon info lengkap kuota sertifikat halal gratis apa syarat perlu”	Informasi dan Prosedur Pendaftaran UMKM	Netral Positif
1	cuan, ladang, baru, tambah, cari, sertifikasi, nyari, pemerintah, oknum, agama, halal, biar, sumber, jadi, lumayan	”inti cari cuan sertifikat cuan”	Kecurigaan Terhadap Oknum dan Komersialisasi	Negatif Skeptis
2	tidak, negara, kalau, jadi, buat, indonesia, halal, apa, udah, pemerintah, bayar, agama, atur, orang, makin	”sertifikat lama cara biar cepat”	Kritik Kebijakan dan Peran Negara	Negatif

Cluster	Kata Dominan	Contoh Komentar	Labeling	Sentimen
3	halal, tidak, non, makan, kalau, label, harus, haram, muslim, sertifikasi, logo, jadi, produk, jual, semua	”tanya serius kalau jual makan non halal label tidak boleh”	Aturan Labeling Produk Nonhalal	Netral Informatif
4	duit, ujung, uud, nyari, ladang, sertifikasi, inti, cari, terus, korupsi, pemerintah, sertifikat, buat, dapat, halal	”sertifikat duit preman alias palak negara”	Isu Korupsi dan Pungutan Liar	Sangat Negatif
5	babi, jual, gimana, kalau, halal, daging, makan, terus, dagang, tidak, non, guling, harus, haram, sertifikat	”kalau dagang babi gimana”	Polemik Produk Nonhalal (Babi)	Netral Diskusi

4.2.6 Distribusi Sentimen

Distribusi sentimen digunakan untuk mengetahui kecenderungan opini masyarakat dari hasil labeling setiap *cluster*. Interpretasi sentimen diberikan ke setiap komentar yang ada pada masing-masing *cluster* sesuai karakteristiknya. Hasil distribusi sentimen akan menampilkan kategori sentimen apa yang paling dominan dalam data komentar. Berikut merupakan grafik distribusi sentimen:



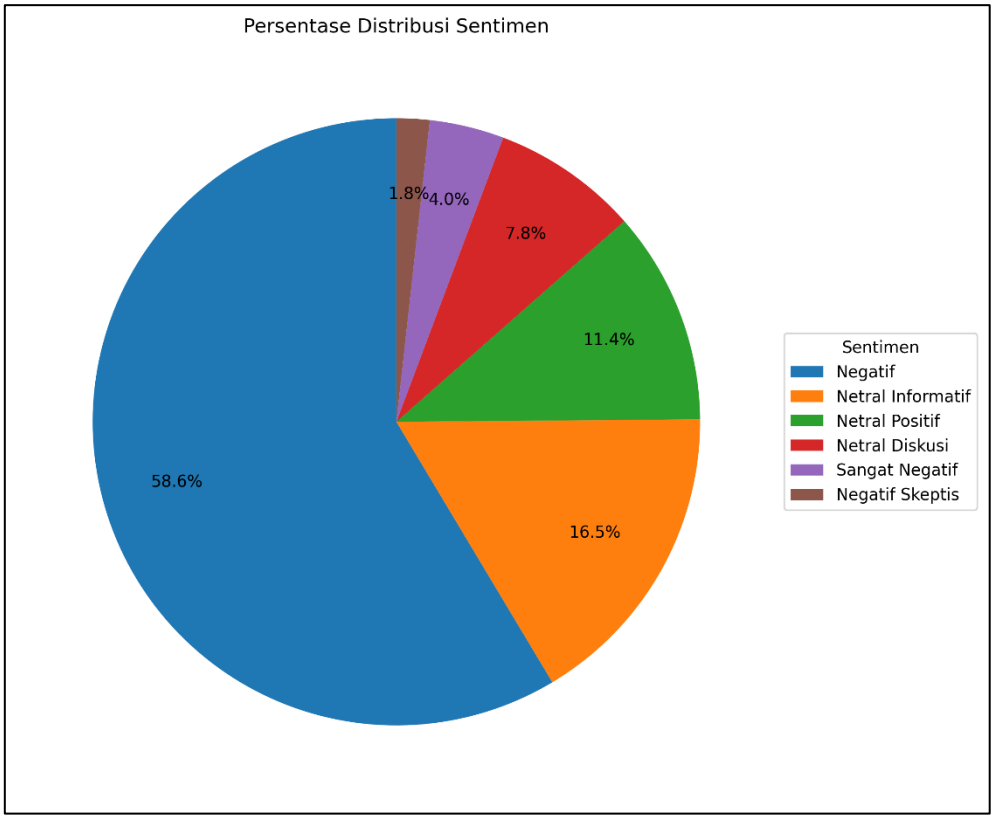
Gambar 4. 9 Grafik Distribusi Sentimen

Berdasarkan grafik tersebut, kategori sentimen negatif paling dominan dengan jumlah 2836 komentar dari keseluruhan data komentar yang berjumlah 4841. Hal ini menunjukkan bahwa sebagian besar opini masyarakat cenderung berisi kritik dan ketidaksetujuan terhadap kebijakan kewajiban sertifikasi halal.

BAB V
PEMBAHASAN

5.1 Deskripsi *Cluster*

Berdasarkan hasil pengolahan data, didapatkan 6 *cluster* optimal yang akan diberikan strategi tertentu sehingga kebijakan kewajiban sertifikasi halal dapat berjalan efektif dan sesuai target. *Cluster* tersebut dikelompokkan berdasarkan kata-kata dominan yang memiliki karakteristik dan pembobotan tertentu. Setiap *cluster* selanjutnya dianalisis lebih dalam untuk menentukan topik apa yang dibahas di dalamnya atau labeling *cluster*. Berikutnya, interpretasi sentimen pada masing-masing *cluster* ditentukan sesuai labelnya. Setelah itu, setiap *cluster* akan dikaji lebih dalam pada profilisasi *cluster* untuk mengetahui karakteristik yang menjadi persamaan dan perbedaan terbentuknya *cluster*.



Gambar 5. 1 Persentase Distribusi Sentimen

Tabel 5. 1 Deskripsi *Cluster*

Cluster	**Kata Dominan**	**Labeling**	**Sentimen**	**Interpretasi**
0	sertifikat, gratis, halal, min, kalau, bayar, buat, tidak, harus, kuota, daftar, usaha, umkm, dapat, sertifikasi	Informasi dan Prosedur Pendaftaran UMKM	Netral Positif	Berfokus pada aspek administratif dan operasional sertifikasi halal bagi pelaku UMKM. Kata kunci seperti gratis, kuota, daftar, dan usaha menunjukkan diskusi mengenai fasilitas sertifikasi halal gratis (SEHATI) dan cara mendapatkannya.
1	cuan, ladang, baru, tambah, cari, sertifikasi, nyari, pemerintah, oknum, agama, halal, biar, sumber, jadi, lumayan	Kecurigaan Terhadap Oknum dan Komersialisasi	Negatif Skeptis	Diskusi mengarah pada pandangan bahwa sertifikasi halal dijadikan sebagai ladang cuan atau sumber pendapatan baru oleh pihak tertentu atau oknum. Terdapat kekhawatiran masyarakat terhadap eksploitasi agama demi keuntungan materi.
2	tidak, negara, kalau, jadi, buat, indonesia, halal, apa, udah, pemerintah, bayar, agama, atur, orang, makin	Kritik Kebijakan dan Peran Negara	Negatif	Berisi opini publik mengenai campur tangan pemerintah/negara dalam mengatur ranah agama melalui aturan hukum. Topik ini berfokus pada ketidaksetujuan atau kritik terhadap kewajiban administratif yang dianggap membebani.
3	halal, tidak, non, makan, kalau, label, harus, haram,	Aturan Labeling	Netral Informatif	Membahas tentang kejelasan identitas produk. Kata kunci seperti label, harus, haram, dan non menunjukkan pentingnya pemisahan yang jelas antara

Cluster	Kata Dominan	Labeling	Sentimen	Interpretasi
	muslim, sertifikasi, logo, jadi, produk, jual, semua	Produk Nonhalal		produk halal dan haram di pasar agar konsumen muslim tidak keliru.
4	duit, ujung, uud, nyari, ladang, sertifikasi, inti, cari, terus, korupsi, pemerintah, sertifikat, buat, dapat, halal	Isu Korupsi dan Pungutan Liar	Sangat Negatif	*Cluster* ini memiliki pembahasan dominan pada kata duit, korupsi, dan ujung-ujungnya duit atau disingkat uud oleh masyarakat. Masyarakat menduga adanya praktik pungli atau korupsi dalam birokrasi pengurusan sertifikat halal.
5	babi, jual, gimana, kalau, halal, daging, makan, terus, dagang, tidak, non, guling, harus, haram, sertifikat	Polemik Produk Nonhalal (Babi)	Netral Diskusi	Secara spesifik membahas komoditas nonhalal seperti babi dan daging haram lain. Diskusi berkaitan dengan bagaimana pedagang makanan nonhalal harus bersikap atau berjualan di tengah aturan sertifikasi yang baru.

Berdasarkan jumlah komentar, *Cluster* 2 memiliki jumlah data terbesar yaitu 2836 komentar atau sebanyak 58,6%. Artinya pembahasan masyarakat mengenai kritik kebijakan dan peran negara menjadi permasalahan yang paling banyak diperbincangkan pada komentar. Komentar pada *Cluster* ini berfokus pada keterlibatan pemerintah dalam mengatur kebijakan sertifikasi halal. Dominasi kata pada komentar yang ada menunjukkan bahwa kebijakan ini dipandang sebagai aturan administratif yang membebani dan peran negara hanya sebagai regulator tanpa memerhatikan beban masyarakat sebagai objek regulasi tersebut.

Cluster 0 berisi 550 komentar yang membahas informasi dan prosedur sertifikasi halal khususnya bagi pelaku UMKM. Kata dominan yang muncul berupa sertifikat, gratis, halal, kuota, daftar, dan usaha menunjukkan bahwa masyarakat membahas banyak akses terhadap program sertifikat halal gratis (SEHATI). *Cluster* ini cenderung memiliki sentimen netral positif karena komentar yang ada berbentuk pertanyaan prosedur dan perhatian terhadap fasilitas yang dapat membantu pelaku usaha kecil dalam memenuhi kewajiban sertifikasi halal.

Cluster 1 memiliki jumlah komentar paling sedikit yaitu 85 komentar. Akan tetapi, *Cluster* ini memiliki karakteristik yang kuat karena terdapat kata dominan seperti cuan,ladang, pemerintah, oknum, agama. Dominasi kata tersebut mengartikan bahwa terdapat kecurigaan masyarakat terhadap kebijakan sertifikasi halal yang dapat dimanfaatkan oleh pihak atau oknum tertentu sebagai sumber keuntungan. Oleh karena itu, *cluster* ini cenderung memiliki sentimen negatif skeptis yang mencerminkan ketidakpercayaan masyarakat dalam pelaksanaan kebijakan.

Cluster 3 dengan jumlah komentar 801 berisi pembahasan masyarakat terhadap aturan labeling pada produk. Kata dominan yang muncul seperti halal, label, haram, makan, produk, logo, sertifikasi mengartikan bahwa masyarakat masih memiliki kebingungan bagaimana perbedaan label halal dan nonhalal dicantumkan pada produk makanan. *Cluster* ini memiliki sentimen netral informatif karena mengandung komentar berupa pertanyaan, penjelasan, serta diskusi terkait aturan kewajiban pencantuman label halal produk.

Cluster 4 berisi 192 komentar dengan kecenderungan sentimen yang sangat negatif. Dominasi kata seperti duit, uud (ujung-ujungnya duit), ladang, korupsi, sertifikasi menunjukkan bahwa sebagian masyarakat berasumsi bahwa kebijakan sertifikasi halal dapat dimanfaatkan sebagai pungutan, biaya tambahan, dan praktik tidak transparan seperti korupsi. Jumlah komentar pada *cluster* ini memang tidak sebanyak *cluster* 2 tetapi karakteristik sentimennya

lebih tajam karena mengarah pada kecurigaan masyarakat langsung kepada pemerintah sebagai penyelenggara kebijakan.

Cluster 5 berjumlah 377 komentar yang berisi posisi produk nonhalal dan sertifikasi. Dominasi kata seperti babi, daging, haram, makan, jual, dan sertifikat dapat diartikan sebagai pertanyaan masyarakat terkait posisi produk nonhalal dalam kebijakan sertifikasi halal. Sentimen yang dimiliki *cluster* 5 cenderung netral bersifat diskusi karena mayoritas komentar dalam bentuk pertanyaan dan perdebatan bagaiman pedagang produk nonhalal harus bersikap ditengah aturan baru ini.

Berdasarkan 6 *cluster* yang telah dibuat, terdapat 3 *cluster* yang memiliki sentimen negatif dan 3 *cluster* lainnya memiliki sentimen netral dan positif. Selanjutnya, pada *cluster* yang memiliki sentimen negatif diberikan strategi dan solusi supaya sentimennya dapat berubah menjadi netral dan positif sebagai salah satu bentuk bahwa sosialisasi dan komunikasi kebijakan sertifikasi halal dapat berhasil. Untuk itu hanya 3 *cluster* yang memiliki sentimen negatif saja yang akan diberikan strategi agar kebijakan sertifikasi halal dapat lebih diterima, dipahami, dan dilakukan oleh masyarakat Indonesia khususnya para pelaku usaha.

5.2 Strategi *Cluster*

Strategi disusun berdasarkan hasil pengelompokkan opini masyarakat pada *cluster* yang optimal. *Cluster* optimal pada penelitian ini merupakan *cluster* yang memiliki sentimen negatif. Sentimen negatif ditunjukkan dengan adanya kritik, rasa khawatir, dan pemahaman yang kurang terhadap kebijakan sertifikasi halal tersebut. Selain itu, kombinasi strategi antar *cluster* negatif dengan *cluster* yang sudah memiliki sentimen netral/positif dapat digabungkan untuk mengarahkan anggota *cluster* negatif menjadi lebih positif. Terdapat tiga *cluster* optimal yang masih memiliki sentimen negatif dan akan diberikan strategi.

Tabel 5. 2 Strategi *Cluster* Optimal

Cluster	**Label**	**Referensi**	**Strategi**
1	Kecurigaan Terhadap Oknum dan Komersialisasi	(Arsyad & Ritonga, 2025); (Ilham et al., 2022); (Sari et al., 2025); (Nofandi & Khalikin, 2025)	Perbaikan komunikasi manfaat sertifikasi halal dengan melibatkan tokoh agama, akademisi, ormas, pemerintah daerah. Mencegah komersialisasi oknum melalui edukasi pendamping halal resmi dan kanal pengaduan
2	Kritik Kebijakan dan Peran Negara	(Nofandi & Khalikin, 2025); (Asmawi & Lutfiadi, 2026); (Setyowati et al., 2024); (Warto et al., 2025); (Hrp et al., 2024); (Lahuri et al., 2025)	Menempatkan pemerintah sebagai fasilitator melalui regulasi yang proporsional, penyederhanaan dokumen, *self declare*, pendampingan, layanan *one stop service*. Menunjukkan manfaat ekonomi yang nyata melalui rekognisi produk halal pada *e-commerce* dan aplikasi layanan pesan antar.
4	Isu Korupsi dan Pungutan Liar	(Simbolon & Hidayat, 2021); (Widodo & Nuryanto, 2024); (Japar et al., 2024); (Faisal & Prabowo, 2025); (Jannah et al., 2025)	Memperkuat transparansi biaya, pembayaran resmi, laporan kinerja, serta tindak lanjut aduan. Menyajikan informasi tersebut dalam bentuk infografik atau konten sederhana sehingga masyarakat dapat dengan mudah memverifikasi pada kanal resmi.

5.2.1 Strategi terhadap *Cluster* Kecurigaan Terhadap Oknum dan Komersialisasi

Cluster 1 menunjukkan pembahasan masyarakat mengarah pada kecurigaan bahwa kebijakan sertifikasi halal ini akan dijadikan sebagai ladang cuan atau sumber pendapatan baru oleh pihak atau oknum tertentu yang memanfaatkan prosedur sertifikasi. Oleh karena itu, strategi pertama yang dapat dilakukan adalah memperbaiki komunikasi publik mengenai tujuan sertifikasi halal. Tidak hanya ditekankan pada kewajiban dan sanksi tetapi dijelaskan sebagai jaminan produk, perlindungan konsumen, peningkatan kepercayaan pasar, dan penguatan daya saing UMKM. Selain itu, beberapa contoh UMKM yang telah berhasil mengembangkan usahanya lebih baik seperti omzet yang meningkat melalui sertifikasi halal perlu dipublikasikan tidak hanya di media sosial tetapi di komunitas UMKM.

Pemerintah juga perlu memperkuat publikasi melalui tokoh agama, akademisi, organisasi masyarakat, pemerintah daerah. Hal ini penting untuk mendapatkan kembali kepercayaan masyarakat karena jika komunikasi hanya dilakukan dan digencarkan oleh pemerintah saja maka sebagian masyarakat akan memandangnya sebagai kepentingan birokratis saja. Melalui publikasi beberapa pihak lain, penerimaan masyarakat terhadap kebijakan menjadi lebih kuat (Arsyad & Ritonga, 2025).

Selanjutnya sebagai upaya pencegahan praktik komersialisasi dan penyalahgunaan prosedur sertifikasi halal oleh oknum tertentu, strategi verifikasi pendamping halal menjadi sangat penting (Ilham et al., 2022). Saat ini sudah terdapat sistem informasi yang membantu masyarakat memeriksa legalitas pendamping halal. Pendamping halal yang kredibel memiliki sertifikat kompetensi pelatihan dan terdaftar dalam sistem resmi aplikasi SIHALAL BPJPH (Sari et al., 2025). Keberadaan daftar resmi pendamping halal (P3H) ini perlu disosialisasikan lebih luas seiring dengan sosialisasi penggunaan aplikasi SIHALAL sehingga pelaku UMKM tidak mudah menggunakan jasa pihak yang tidak terverifikasi. Setiap sosialisasi dan kampanye sertifikasi halal sebaiknya tidak hanya menjelaskan cara mendaftar tetapi juga menegaskan dan mengajarkan pelaku usaha untuk mengecek pendamping resmi, mengenali tanda-tanda jasa tidak resmi, dan melaporkan permintaan biaya di luar ketentuan. Pelaku usaha perlu diberikan informasi bahwa pendamping resmi harus memiliki identitas kelembagaan yang jelas, tidak meminta pembayaran pribadi di luar ketentuan, dan menggunakan aplikasi resmi dalam proses pengajuan terutama untuk produk UMK yang menggunakan program *self declare*.

Apabila terdapat permintaan persyaratan di luar ketentuan resmi maka hal tersebut harus dapat dilaporkan melalui pengaduan yang mudah diakses seperti WhatsApp resmi dan pelayanan daerah setempat. Pemerintah perlu menindak tegas terhadap segala bentuk pungutan liar dan jasa percepatan proses sertifikasi halal. Strategi tersebut sejalan dengan penelitian (Nofandi & Khalikin, 2025) yang menegaskan hambatan sertifikasi halal berkaitan dengan aspek tata kelola seperti mekanisme pengaduan, evaluasi, dan pengawasan.

5.2.2 Strategi terhadap *Cluster* Kritik Kebijakan dan Peran Negara

Cluster 2 adalah kelompok yang memiliki anggota terbesar dalam penelitian ini, opini masyarakat dalam *cluster* ini perlu diarahkan bahwa peran negara bukan sekadar regulator tetapi fasilitator kepatuhan. Dominasi komentar negatif pada *cluster* ini berisi keterlibatan negara dan beban biaya sehingga pemerintah perlu mengedepankan fungsi fasilitasi optimal dengan menjamin tersedianya edukasi, pendampingan, dan masa transisi sebelum diberlakukan sanksi secara keseluruhan. Hal ini akan mengedepankan pelaku usaha menjadi subjek kebijakan yang diberdayakan bukan objek regulasi yang sejalan dengan penelitian (Nofandi & Khalikin, 2025) berisi penegasan bahwa efektivitas sertifikasi bergantung pada kesiapan tata kelola serta dukungan penuh terhadap kapasitas UMKM pada era digital.

Pemerintah juga perlu menerapkan regulasi sesuai dengan proporsi setiap skala usaha dan tingkat risiko produk. Perbedaan standar antara pelaku usaha besar yang memiliki kapasitas finansial lebih dari UMKM mikro sangat penting. Pada sektor pelaku usaha mikro dan rumah tangga, pemerintah perlu menyampaikan jelas dan menggencarkan penyederhanaan dokumen serta penguatan skema *self declare* bagi produk berisiko rendah sehingga opini masyarakat terkait beban administratif dan proses rumit akan berubah menjadi pemahaman bahwa proses sertifikasi mudah dan bagi masyarakat yang sudah yakin proses produksinya halal akan lebih percaya dan menerima skema *self declare* ini tanpa merasa terbebani. Strategi ini relevan dengan penelitian (Asmawi & Lutfiadi, 2026) yang menekankan regulasi proporsional adalah kunci mengurangi hambatan finansial dan fasilitas dalam mengakses legalitas halal.

Kemudian, strategi tersebut diperkuat juga melalui layanan *one stop service* yang terintegrasi baik secara *online* maupun *offline*. Layanan *offline* dapat diadakan di desa, kecamatan, pasar rakyat, sentra UMKM, atau dinas daerah setempat sebagai penghubung pelaku usaha yang belum mampu menggunakan layanan digital secara mandiri dan layanan dapat diakses di berbagai daerah. Layanan tersebut akan dibantu oleh pendamping halal dalam mengurus NIB melalui OSS RBA, membuat akun pada aplikasi SIHALAL, mengecek kriteria

program SEHATI dan *self declare*, menyiapkan dokumen, mengunggah data, serta memantau status pengajuan. Selain itu, pemerintah perlu membuat peta alur layanan sertifikasi halal nasional yang mampu menjelaskan hubungan antara OSS RBA, aplikasi SIHALAL, SEHATI, PTSP halal, pendamping halal beserta lembaga terkait lainnya. Peta alur tersebut harus dibagikan tidak hanya melalui media sosial tetapi juga *website* resmi, kantor desa, kecamatan, dinas UMKM dan koperasi, pasar rakyat, serta komunitas UMKM sehingga pelaku usaha dapat memahami proses sertifikasi halal dengan lebih sederhana. Penelitian (Setyowati et al., 2024) menunjukkan bahwa pendampingan langsung penggunaan aplikasi SIHALAL membantu kelancaran proses sertifikasi halal. Strategi ini akan membuat digitalisasi tidak hanya dijangkau pelaku usaha yang sudah paham teknologi tetapi juga bagi seluruh lapisan pelaku usaha mikro dan kecil sehingga rendahnya literasi digital sebagai hambatan proses sertifikasi halal dapat diatasi.

Selain penyederhanaan layanan melalui peta alur, pemerintah dapat menambah jumlah pendamping proses produk halal secara merata di setiap daerah. Penambahan jumlah pendamping ini penting karena proses sertifikasi halal bagi UMKM tidak hanya membutuhkan aplikasi, diperlukan bantuan untuk memahami persyaratan seperti dokumen tertentu, mengidentifikasi bahan baku, membuat alur produksi, mengunggah data pada aplikasi SIHALAL, dan memastikan keseluruhan proses produksi sesuai dengan ketentuan halal. Jika jumlah pendamping halal tidak sebanding dengan jumlah UMKM yang ditargetkan untuk sertifikasi halal pada setiap daerah maka pemberlakuan sertifikasi halal 2026 untuk keseluruhan pelaku UMKM terutama makanan akan berpotensi mengalami keterlambatan. Strategi ini sesuai dengan penelitian (Warto et al., 2025) dan (Hrp et al., 2024) yang menemukan bahwa sebagian UMKM masih memiliki waktu tunggu panjang pada beberapa daerah yang kekurangan pendamping halal sehingga pendamping halal perlu ditingkatkan agar dapat memperluas sertifikasi halal mandiri bagi pelaku UMKM.

Penambahan jumlah pendamping halal dapat dilakukan melalui kerja sama antara pemerintah daerah dengan perguruan tinggi, organisasi masyarakat, dinas UMKM dan koperasi, komunitas usaha. Penambahan pendamping halal ini juga perlu ditunjang dengan peningkatan kualitas layanan seperti pelatihan berstandar, pembagian wilayah kerja yang jelas, sistem evaluasi kinerja, insentif yang memadai, kanal koordinasi antara pendamping dengan BPJPH, pemerintah daerah, dan pelaku UMKM. Selanjutnya, sebaran kuota untuk program SEHATI dan ketersediaan kuota yang masih ada untuk setiap daerah perlu ditampilkan pada

interface aplikasi layanan SIHALAL sehingga pelaku usaha dapat mengecek sendiri apakah kuota masih tersedia atau tidak.

Berikutnya, agar sertifikat halal tidak dipandang sebagai beban administratif saja, pemerintah dapat mendukung skema sertifikasi halal *self declare* dengan integrasi sistem digital yang *user friendly* mencakup pemberian insentif pasar seperti kerja sama dengan *platform e-commerce* untuk memberikan rekognisi atau pengakuan khusus bagi produk yang telah memiliki sertifikat halal. Dengan demikian, selain kesederhanaan regulasi dan biaya yang lebih murah, skema *self declare* akan menjadi kebutuhan strategis UMK dalam meningkatkan nilai tambah dan daya saing produk di pasar nasional maupun global. Hal ini sesuai dengan penelitian (Lahuri et al., 2025) yang menyampaikan bahwa tata kelola sertifikasi halal harus berpijak pada kemanfaatan publik yang nyata tanpa mengabaikan kenyataan pada aspek sosial ekonomi masyarakat Indonesia.

5.2.3 Strategi terhadap *Cluster* Dugaan Pungutan dan Korupsi

Cluster 4 menunjukkan bahwa kecurigaan masyarakat tidak hanya tertuju pada oknum/calo atau pendamping sertifikasi seperti di *cluster* 2 tetapi langsung kepada pemerintah sebagai penyelenggara kebijakan. Oleh karena itu, diperlukan strategi yang berfokus pada pembuktian akuntabilitas pemerintah. Strategi utama yang dapat dilaksanakan adalah menyampaikan secara rinci dasar penetapan biaya sertifikasi halal untuk jenis *self declare* dan reguler. Pemerintah harus menjelaskan komponen biaya, pihak yang menerima pembayaran, alasan biaya dibebankan kepada pelaku usaha sehingga masyarakat mengetahui bahwa terdapat aturan jelas mengenai perincian biaya sertifikasi halal. Hal ini penting karena biaya sertifikasi termasuk salah satu sumber yang membuat masyarakat dan pelaku usaha keberatan. Penelitian (Simbolon & Hidayat, 2021) menjelaskan bahwa proses sertifikasi halal di Indonesia masih memerlukan persoalan biaya yang transparan dan akuntabel sehingga tidak menimbulkan kecurigaan ataupun ketidakjelasan pada prosesnya.

Selanjutnya, diperlukan laporan kinerja beserta penggunaan anggaran sertifikasi halal secara berkala yang dipublikasikan transparan. Dalam praktiknya, BPJPH sebenarnya telah menyediakan sejumlah informasi publik seperti akuntabilitas yang berisikan laporan kinerja dan hasil evaluasi sertifikasi halal setiap tahunnya. Oleh karena itu, strategi yang perlu dilakukan bukan hanya menyediakan data tetapi memastikan masyarakat mengetahui dan mampu memahami informasi yang telah disediakan sebelum memberikan opini negatif terhadap pemerintah. BPJPH perlu mengubah informasi yang telah tersedia di *website* resmi

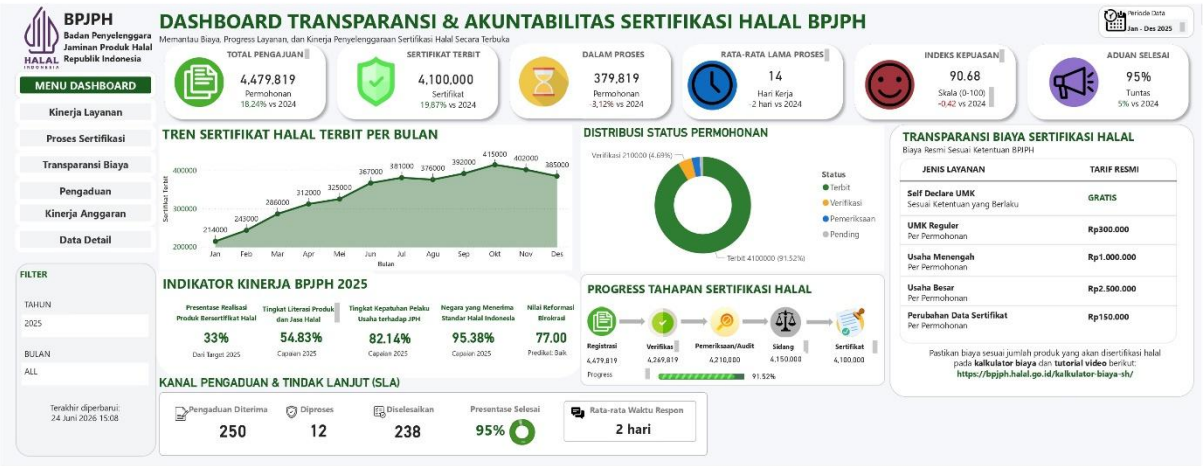
menjadi bentuk komunikasi yang lebih mudah dipahami seperti infografik uang sertifikasi halal digunakan untuk apa, video singkat cara menggunakan kalkulator biaya yang terdapat di *website* sebagai estimasi biaya yang akan dikeluarkan pelaku usaha, serta penjelasan sederhana mengenai isi laporan akuntabilitas dan tindak lanjut aduan. Hal ini penting karena sebagian masyarakat yang berkomentar negatif mungkin tidak membaca dokumen resmi dan tidak memahami istilah administratif yang tersedia di *website*. Penyajian yang lebih mudah dipahami secara visual akan memudahkan masyarakat mengerti bahwa biaya sertifikasi halal memiliki dasar perhitungan dan mekanisme pertanggungjawaban yang jelas. Strategi ini relevan terdapat penelitian (Widodo & Nuryanto, 2024) yang menekankan pengawasan jaminan produk halal BPJPH memerlukan dukungan transparansi serta efisiensi sehingga dapat membangun kepercayaan publik.

Berikutnya, untuk mengatasi narasi masyarakat terkait kata dominan ”ujung-ujungnya duit” dapat dilakukan dengan membuat konten berbasis data visual seperti membandingkan kalkulator biaya, laporan akuntabilitas, dan rekapitulasi aduan menggunakan bahasa yang edukatif sehingga masyarakat dapat diarahkan untuk melakukan verifikasi informasi melalui kanal resmi *website* BPJPH bukan hanya opini media sosial semata. Hal ini sejalan dengan penelitian (Japar et al., 2024) bahwa lembaga sertifikasi halal turut bertanggung jawab dalam menjaga kepercayaan publik dalam ekosistem halal. Kemudian, segala informasi mengenai transparansi biaya tersebut dapat disebarkan melalui media sosial secara berkala sehingga dapat dijangkau oleh seluruh kelompok masyarakat terutama yang menjadi target kebijakan sertifikasi halal.

Selain evaluasi dan transparansi anggaran, diperlukan evaluasi berkala penyerapan kuota sertifikasi halal gratis (SEHATI). Jika hasil evaluasi menunjukkan bahwa ada sisa kuota yang tidak terserap dalam jangka waktu tertentu maka dapat dilakukan mekanisme realokasi kuota secara hati-hati kepada daerah dengan kebutuhan kuota lebih tinggi. Akan tetapi, diperlukan kehati-hatian tinggi agar realokasi kuota tidak merugikan daerah lain dengan akses layanan rendah. Realokasi kuota tidak boleh dilakukan hanya berdasarkan rendahnya angka pemakaian kuota di suatu daerah tetapi harus mempertimbangkan jumlah pelaku UMK daerah tersebut, kesiapan dan jumlah pendamping halal yang tersedia di daerah setempat, dan tingkat permohonan sertifikasi halal sejalan dengan penelitian (Faisal & Prabowo, 2025) yang menunjukkan bahwa pelaksanaan program SEHATI bergantung pada kesiapan pelaksana lapangan termasuk adanya pendamping halal yang aktif. Berikutnya (Jannah et al., 2025) juga

menyampaikan bahwa kemudahan akses program SEHATI harus dilihat dalam kerangka keadilan distributif. Hal ini karena tidak semua daerah memiliki tingkat akses informasi dan kesiapan administrasi yang sama.

Kemudian, pada menu tampilan literasi di *website* resmi BPJPH, konten literasi dapat diperbarui secara berkala karena sekarang masih memuat artikel lama tahun 2024. Isi dari menu literasi perlu disesuaikan juga dengan isu aktual yang relevan dan banyak berkembang di masyarakat, bukan hanya literasi mengenai prosedur sertifikasi halal saja. Misalnya berdasarkan permasalahan di *cluster* 4 ini, konten literasi dapat diarahkan untuk menjawab isu sertifikasi halal sebagai ladang uang pemerintah. Dengan demikian, literasi tidak hanya informatif tetapi juga mengoreksi persepsi negatif masyarakat. Berikut usulan *design engineering* berupa *dashboard* transparansi akuntabilitas dari laporan tranpasaransi kinerja BPJPH menggunakan aplikasi Power BI. Dashboard ini dapat memudahkan masyarakat mengakses informasi transparansi biaya, pengajuan sertifikat per bulannya, dan progress pengaduan pelaku usaha yang muncul dan telah diselesaikan BPJPH.



Gambar 5. 2 *Design Engineering* Dashboard Transparansi Akuntabilitas BPJPH

Sebagai bentuk implementasi strategi pada *cluster* 4 yang membahas mengenai dugaan korupsi pada kebijakan sertifikasi halal oleh pemerintah, penelitian *mengusulkan desain engineering* yang dirancang menggunakan Power BI berupa *dashboard* transparansi akuntabilitas sertifikasi halal. *Dashboard* ini berfungsi sebagai media pemantauan untuk masyarakat dan meningkatkan literasinya pada kebijakan sertifikasi halal ini dengan informasi yang mudah diakses dan dipahami. *Dashboard* tersebut memiliki beberapa komponen utama. Pertama, ada indikator total pengajuan dan sertifikat terbit untuk memantau jumlah sertifikat

yang sudah berhasil terbit serta sebagai bahan evaluasi jika terdapat ketidaksesuaian antara jumlah permohonan dengan sertifikat yang diterbitkan. Kedua, terdapat indikator permohonan dalam proses serta rata-rata lama proses yang berfungsi untuk memantau efisiensi pelayanan. Informasi yang disajikan adalah jumlah permohonan yang sedang diproses dan waktu proses sehingga dapat mengidentifikasi jika ada keterlambatan pelayanan. Ketiga, terdapat informasi biaya resmi sertifikasi halal yang terhubung dengan kalkulator biaya sertifikat halal yang sudah ada pada *website* resmi BPJPH sehingga masyarakat memiliki kepastian biaya resmi tanpa khawatir akan ada biaya pungutan lain.

Selanjutnya, *dashboard* juga memiliki indikator kinerja BPJPH berupa tingkat realisasi produk bersertifikat halal, tingkat literasi pelaku usaha, tingkat kepatuhan pelaku usaha terhadap Jaminan Produk Halal (JPH), jumlah negara yang menerima standar halal Indonesia, dan nilai reformasi birokrasi yang menggambarkan capaian kinerja dan efektivitas penyelenggaraan kebijakan. Terdapat juga alur proses pengajuan sertifikasi halal secara visual yang dapat dipahami masyarakat luas. Selain itu, *dashboard* menampilkan kanal pengaduan dan tindak lanjutnya. Indikatornya berupa jumlah pengaduan yang diterima, diproses, telah diselesaikan, serta presentase penyelesaiannya setiap laporan yang diterima. *Dashboard* transparansi akuntabilitas ini diharapkan dapat menjadi pendukung proses pengambilan keputusan, meningkatkan keterbukaan informasi dan literasi masyarakat, dan mempermudah masyarakat mengakses informasi terkait akuntabilitas kebijakan kewajiban sertifikasi halal ini sehingga kepercayaan masyarakat terhadap pelaksanaan kebijakan ini dapat terus meningkat.

5.3 Limitasi Penelitian

Berdasarkan hasil penelitian yang diperoleh, terdapat beberapa limitasi penelitian. Limitasi ini tidak mengurangi nilai penelitian tetapi menjadi batasan pada proses pengumpulan data, pengolahan data, interpretasi hasil, dan perumusan strategi. Penelitian ini hanya menggunakan satu algoritma, yaitu *K-Means Clustering*. Hasil pengelompokkan bergantung pada karakteristik *K-Means* dimana algoritma ini memiliki kelemahan karena sensitif terhadap pemilihan jumlah *cluster*. Selain itu, metode *Elbow* yang digunakan juga bersifat subjektif karena titik siku ditentukan berdasarkan interpretasi visual pada grafik WCSS.

Kemudian, penelitian ini hanya menggunakan data komentar yang bersumber dari YouTube dan Instagram dimana penggunaan kedua *platform* tersebut belum tentu dapat merepresentasikan seluruh opini masyarakat Indonesia. Penelitian ini juga hanya menganalisis data berupa teks komentar yang telah melalui preprocessing dan pembobotan TF-IDF dimana

penelitian ini belum mempertimbangkan aspek lain seperti gambar, emoji, dan intonasi yang mungkin dapat memengaruhi makna komentar.

Selain itu, proses labeling dan interpretasi sentimen masih melibatkan interpretasi peneliti walaupun proses tersebut didasarkan pada pembobotan kata dominan sehingga perlu dipahami bahwa interpretasi belum melibatkan validasi dari ahli. Selanjutnya, strategi dan rekomendasi yang diusulkan masih bersifat rekomendasi dan belum diuji secara langsung kepada pihak terkait seperti BPJPH, pemerintah daerah, auditor halal, pendamping halal, pelaku UMKM maupun masyarakat sebagai penerima kebijakan. Oleh karena itu, strategi yang diusulkan dapat digunakan sebagai rekomendasi awal yang perlu diuji dan dikembangkan lebih lanjut pada penelitian selanjutnya.

BAB VI
PENUTUP

6.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan sebelumnya, berikut merupakan kesimpulan:

1. Kecenderungan sentimen masyarakat Indonesia terhadap kebijakan kewajiban sertifikasi halal pada komentar YouTube dan Instagram didominasi oleh sentimen negatif, terutama pada *cluster* 2 yaitu sejumlah 2836 komentar atau 58,6% dari keseluruhan data komentar. Terdapat pula *Cluster* 1 yang memiliki sentimen negatif skeptis serta *cluster* 4 dengan sentimen sangat negatif. Interpretasi sentimen tersebut didasarkan pada pengelompokkan pola opini yang berasal dari anggota kata dominan pada setiap *cluster*. Hal ini menunjukkan bahwa penerimaan masyarakat terhadap kebijakan kewajiban sertifikasi halal masih menghadapi tantangan.
2. Pola opini masyarakat terhadap kebijakan kewajiban sertifikasi halal terbagi dalam enam *cluster* yang memiliki karakteristik pembahasan berbeda. *Cluster* 0 berisi topik bahasan mengenai informasi dan prosedur pendaftaran sertifikasi halal terutama pada program SEHATI, tersedianya kuota sertifikat gratis, dan cara pendaftaran. *Cluster* 1 berisi kecurigaan masyarakat terhadap pemanfaatan keuntungan oleh oknum tertentu. *Cluster* 2 berisi kritik terhadap kebijakan dan peran negara yang dipandang terlalu mengatur dan menambah beban administratif untuk masyarakat dan pelaku usaha. *Cluster* 3 berisi diskusi aturan labeling produk nonhalal. *Cluster* 4 membahas dugaan masyarakat terhadap adanya pungutan liar, biaya tambahan, dan korupsi dalam proses sertifikasi halal yang dilakukan oleh pemerintah. *Cluster* 5 berisi polemik produk nonhalal seperti babi dan makanan haram lainnya mengenai posisi dan sikap pedagang produk nonhalal dalam penerapan kebijakan sertifikasi halal.
3. Penyusunan strategi difokuskan pada *cluster* yang masih memiliki sentimen negatif, yaitu *cluster* 1, *cluster* 2, dan *cluster* 4. *Cluster* 1 memiliki strategi perbaikan komunikasi manfaat sertifikasi halal, melibatkan pihak terpercaya seperti tokoh agama, akademisi, organisasi masyarakat, dan pemerintah daerah bersamaan dengan melakukan edukasi pendamping halal resmi untuk mencegah komersialisasi yang dapat dilakukan oleh oknum tertentu. Pada *cluster* 2, pemerintah perlu diarahkan sebagai fasilitator dengan regulasi yang proporsional, penyederhanaan dokumen, penguatan skema *self declare*,

layanan *one stop service*, penambahan jumlah pendamping halal, serta pemberian manfaat ekonomi seperti rekognisi produk halal di *e-commerce* atau aplikasi layanan pesan antar. Dalam *cluster* 4, strategi dapat berupa transparansi biaya, pembayaran resmi, laporan kinerja, tindak lanjut pengaduan, serta penyajian informasi dalam bentuk infografik atau konten yang lebih sederhana secara visual sehingga masyarakat lebih mudah memahami dan melakukan verifikasi informasi melalui kanal resmi.

6.2 Saran

Berdasarkan limitasi penelitian yang telah dijelaskan sebelumnya, berikut merupakan beberapa saran yang dapat diberikan untuk penelitian selanjutnya:

1. Penelitian selanjutnya dapat membandingkan hasil dari algoritma *K-Means Clustering* dengan algoritma lain seperti *hierarchical clustering*, DBSCAN, atau algoritma supervised seperti SVM, *naïve bayes*, *random forest*, dan IndoBERT. Perbandingan metode tersebut dapat membantu mengetahui konsistensi pola sentimen dan opini yang terbentuk.
2. Penelitian selanjutnya dapat menambahkan sumber data dari *platform* media sosial lainnya seperti X, TikTok, Facebook, portal berita *online*, forum komunitas. Penambahan sumber data dapat memberikan gambaran opini masyarakat yang lebih luas dan seragam sehingga dapat sepenuhnya merepresentasikan opini seluruh masyarakat Indonesia terhadap kebijakan kewajiban sertifikasi halal.
3. Penelitian selanjutnya dapat mempertimbangkan konten selain data teks komentar, seperti gambar, emoji, intonasi, dan hubungan antar komentar menggunakan pendekatan analisis kontekstual sehingga makna komentar dapat dipahami lebih dalam.
4. Penelitian selanjutnya dapat melibatkan teknik validasi lain selain dari referensi jurnal, seperti meminta pendapat ahli seperti auditor halal dan lembaga pendamping proses produk halal (LP3H) menggunakan metode delphi, FGD, dan analisis kelayakan strategi.

DAFTAR PUSTAKA

Afandi, W., Laksana, T. G., Annisa, N., & Tanjung, F. (2023). *Penerapan Metode Support Vector Machine Analisis Sentimen Tweet Pergantian Logo Halal di Indonesia*. *16*(1), 45–52.

Aggarwal, C. C., & Zhai, C. (2012). *An Introduction to Text Mining BT - Mining Text Data* (C. C. Aggarwal & C. Zhai (eds.); pp. 1–10). Springer US. https://doi.org/10.1007/978-1-4614-3223-4_1

Aji Nursalim, & Rice Novita. (2023). SENTIMENT ANALYSIS OF COMMENTS ON GOOGLE PLAY STORE, TWITTER AND YOUTUBE TO THE MYPERTAMINA APPLICATION WITH SUPPORT VECTOR MACHINE. *Jurnal Teknik Informatika (Jutif)*, *4*(6 SE-Articles), 1305–1312. https://doi.org/10.52436/1.jutif.2023.4.6.1059

AL-Sharuee, M. T., Liu, F., & Pratama, M. (2018). Sentiment analysis: An automatic contextual analysis and ensemble clustering approach and comparison. *Data & Knowledge Engineering*, *115*, 194–213. https://doi.org/https://doi.org/10.1016/j.datak.2018.04.001

Ariel, Q., Chang, V., & Jayne, C. (2022). *A systematic review of social media-based sentiment analysis : Emerging trends and challenges ☆*. *3*(April).

Arsyad, A., & Ritonga, Y. E. (2025). Strategi Komunikasi Dakwah Badan Penyelenggara Jaminan Produk Halal Dalam Edukasi Sertifikasi Halal Pada UMKM Kuliner Kota Medan. *Jurnal Khabar: Komunikasi Dan Penyiaran Islam*, *7*(1), 339–352. https://doi.org/https://doi.org/10.37092/khabar.v7i1.1172

Asmawi, M., & Lutfiadi, A. (2026). *Challenges Implementing Indonesia ' s Halal Product Guarantee Law for MSMEs : Legal Readiness Review*. *1*(1), 125–142.

Badan Penyelenggara Jaminan Produk Halal. (2026). *Kabar Gembira, BPJPH Buka Kuota 1,35 Juta Sertifikasi Halal Gratis 2026 bagi UMK*. https://bpjph.halal.go.id/detail/kabar-gembira-bpjph-buka-kuota-1-35-juta-sertifikasi-halal-gratis-2026-bagi-umk-1

Baker, L. (2022). *Data Cleaning: The Ultimate Practical Guide: From Dirty Data to Clean Data*. Lee Baker. https://books.google.co.id/books?id=TJCaEAAAQBAJ

Barua, T., Hiran, K. K., Jain, R. K., & Doshi, R. (2024). *Machine Learning with Python*. De Gruyter. https://books.google.co.id/books?id=fSU7EQAAQBAJ

Bimantoro, F. (2025). *Analisis Sentimen Ulasan Wisatawan Terhadap Destinasi Gili di KLU Menggunakan K-Means Clustering*. *7*(2), 177–189.

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134. https://doi.org/https://doi.org/10.1016/j.knosys.2021.107134

Bordoloi, M., & Biswas, S. K. (2023). Sentiment analysis: A survey on design framework, applications and future scopes. *Artificial Intelligence Review*, *56*(11), 12505–12560. https://doi.org/10.1007/s10462-023-10442-2

CNBC Indonesia. (2026, February 23). Wajib Halal Oktober, BPJPH Sebut Baru 35% UMKM

Tersertifikasi. *CNBC Indonesia*. https://www.cnbcindonesia.com/news/20260223122620-4-713024/wajib-halal-oktober-bpjph-sebut-baru-35-umkm-tersertifikasi

DataReportal. (2025). *Digital 2025: Indonesia*. https://datareportal.com/reports/digital-2025-indonesia

Denha, B., Ummah, S., Rahmawati, A., Ridho, A., & Harahap, A. A. (2024). Assessing the Legal Impact of Halal Certification Obligations Through the Lens of Social Awareness and Consumer Rights Protection. *Jurnal Indo Islamika*, *14*(2), 393–407.

Djamaluddin, M., Darajat, D. M., & Rahmi, C. (2024). *Strategi Komunikasi Publik Badan Penyelenggara Jaminan Produk Halal dalam Sosialisasi Sertifikasi Halal di Media Sosial*. *28*(1), 13–24.

Faisal, F., & Prabowo, S. (2025). Evaluasi Pelaksanaan Program Sertifikasi Halal Gratis (SEHATI) 2022 di Kecamatan Muara Jawa, Kutai Kartanegara-Kalimantan Timur. *Jurnal Ilmiah Pangan Halal*, *7*(2), 162–169.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.

Fernandes, A. A. R., Erwinda, E. G., Hardianti, R., & Arini, L. H. Y. (2023). *Metodologi Penelitian: Variabel Mining berbasis Big Data dalam Pemodelan Sistem untuk mengungkap Research Novelty*. Universitas Brawijaya Press. https://books.google.co.id/books?id=GfXaEAAAQBAJ

Geetha, T. V, & Sendhilkumar, S. (2023). *Machine Learning: Concepts, Techniques and Applications*. CRC Press. https://books.google.co.id/books?id=iKEIEQAAQBAJ

Harahap, M. A., & Ikhsan, M. (2024). *Analysis of Public Opinion Sentiment towards the 2024 Presidential Election Based on Clustering Method with K - Means Algorithm*. *4*(2), 594–608.

Hickman, Louis, Thapa, Stuti, Tay, Louis, Cao, Mengyang, & Srinivasan, Padmini. (2022). Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations. *Organizational Research Methods*, *25*(1), 114–146. https://doi.org/10.1177/1094428120971683

Hikmah, N., Fadillah, A., Intan, S. M., Yunus, M., Islam, U., & Utara, S. (2025). *Tantangan Sertifikasi Halal pada UMKM di Indonesia : Studi Kualitatif terhadap Aspek Sosialisasi , Biaya , dan Regulasi*. *5*.

Hrp, A. K. Z., Fitri, S., & Batubara, Y. (2024). Pendampingan Sertifikasi Halal Mandiri (Self Declare) Pada UMKM Kabupaten Mandailing Natal. *BUDIMAS: JURNAL PENGABDIAN MASYARAKAT*, *6*(1).

Ibrahim, C., Jaya, A., Simatupang, Y., Daado, J., Azizah, N., Ismah, F., & Ansar, F. (2021). *THE SENTIMENT ANALYSIS OF INDONESIAN NATIONAL LIBRARY ' S TWITTER AND INSTAGRAM*. *5*(2), 48–56.

Ilham, B. U., Sultan, J., No, A., & Pos, K. (2022). *Pendampingan Sertifikasi Halal Self Declare pada Usaha Mikro dan Kecil Binaan Pusat Layanan Usaha Terpadu Sulawesi Selatan*. *05*, 20–25.

Indonesian Chamber of Commerce and Industry. (2024). *UMKM Indonesia*. https://kadin.id/data-dan-statistik/umkm-indonesia/

Jannah, I., Ariswanto, D., & Setiawan, F. (2025). Aksesibilitas Program Sertifikasi Halal Gratis (SEHATI) 2025 di Kabupaten Bangkalan dalam Kerangka Keadilan Distributif. *Al-Maktabah: Jurnal Studi Islam Interdisiplin*, *2*(2), 75–89. https://journal.syamilahpublishing.com/maktabah/article/view/514

Japar, R., Paraikkasi, I., & Muthiadin, C. (2024). PERAN LEMBAGA SERTIFIKASI HALAL DALAM MEMBANGUN EKOSISTEM HALAL: TANTANGAN DAN PELUANG. *International Journal Mathla'ul Anwar of Halal Issues*, *4*(2), 34–44. https://doi.org/https://doi.org/10.30653/ijma.202442.111

Jardim, S., Mora, C., Jardim, S., & Mora, C. (2022). Customer reviews sentiment-based analysis and clustering for tourism services and products development or Customer reviews and clustering for positioning market-oriented tourism services and products development or positioning. *Procedia Computer Science*, *196*(2021), 199–206. https://doi.org/10.1016/j.procs.2021.12.006

Kabir, M. E. (2022). *Topic and sentiment analysis of responses to Muslim clerics ' misinformation correction about COVID-19 vaccine : Comparison of three machine learning models*. *1*(3), 497–523.

Kaushik, N., & Bhatia, M. K. (2022). *Twitter Sentiment Analysis Using K-means and Hierarchical Clustering on COVID Pandemic BT - International Conference on Innovative Computing and Communications* (A. Khanna, D. Gupta, S. Bhattacharyya, A. E. Hassanien, S. Anand, & A. Jaiswal (eds.); pp. 757–769). Springer Singapore.

Kawsar, N. H. (2025). *State of the Global Islamic Economy*.

Kementerian Dalam Negeri. (2026). Satu Data Seribu Wajah, Kemendagri Ungkap Keragaman Indonesia Lewat Data Kependudukan. *Kementerian Dalam Negeri*. https://dukcapil.kemendagri.go.id/blog/read/satu-data-seribu-wajah-kemendagri-ungkap-keragaman-indonesia-lewat-data-kependudukan

Kim, J. Y., Fusaroli, R., Park, H. W., & Bechmann, A. (2021). *Systemic Gendering in Facebook Group Participation*. https://doi.org/10.1177/20563051211064906

Korpela, J. K. (2006). *Unicode Explained*. O'Reilly Media. https://books.google.co.id/books?id=lxndiWaFMvMC

Lahuri, S. Bin, Pranoto, M., Kamaluddin, I., Syamsuri, S., Cahyo, E., & Yusuf, M. (2025). Halal Certification Governance in Indonesia: An Islamic Legal Appraisal Based on Maṣlaḥah Principles. *JURIS (Jurnal Ilmiah Syariah)*, *24*, 233–247. https://doi.org/10.31958/juris.v24i2.15411

Lee, G., Kim, C. O., & Song, M. (2020). Semisupervised sentiment analysis method for online text reviews. *Journal of Information Science*, *47*, 016555152091003. https://doi.org/10.1177/0165551520910032

Lestari, S., & Febrianti, S. (2025). *Sentiment Analysis of Shopee Product Reviews on the Instagram Application Using the K-Nearest Neighbors Algorithm Analisis Sentimen Ulasan Produk Shopee di Aplikasi Instagram*. *5*(October), 1172–1180.

Liu, B. (2012). *Sentiment Analysis: A Fascinating Problem BT - Sentiment Analysis and Opinion Mining* (B. Liu (ed.); pp. 1–8). Springer International Publishing. https://doi.org/10.1007/978-3-031-02145-9_1

LLC, C. T. (2025). *Natural Language Processing with Python: Master text processing, language modeling, and NLP applications with Python's powerful tools*. Packt Publishing. https://books.google.co.id/books?id=EG4_EQAAQBAJ

Melati, R., & Islami, M. R. R. (2024). *ANALISIS SENTIMEN DATA TWITTER MENGGUNAKAN METODE K-MEANS CLUSTERING PADA STUDI KASUS PEMINDAHAN IBU KOTA NUSANTARA (IKN)*. *15*(1), 66–73.

Moscato, R., Ligios, A., & Silvestri, C. (2025). *Natural Language Processing Cookbook: Step-by-step practical solution for unlocking the power of natural language processing potential (English Edition)*. Walter de Gruyter GmbH. https://books.google.co.id/books?id=j_JJEQAAQBAJ

Muhariya, A., Riadi, I., Prayudi, Y., & Saputro, I. A. (2023). *Ingénierie des Systèmes d ' Information Utilizing K-means Clustering for the Detection of Cyberbullying Within Instagram Comments*. *28*(4), 939–949.

Murthy, D. (2024). *Sociology of Twitter / X : Trends , Challenges , and Future Research Directions*. 169–190.

Nofandi, R. A., & Khalikin, A. (2025). *BARRIERS TO HALAL CERTIFICATION AMONG MICRO , SMALL , AND MEDIUM ENTERPRISES IN MAKASSAR : FINANCIAL , DIGITAL , AND GOVERNANCE*. *25*(2). https://doi.org/10.3390/su14042152.2

Noring, C., Jain, A., Fernandez, M., Mutlu, A., & Jaokar, A. (2024). *AI-Assisted Programming for Web and Machine Learning*. Packt Publishing. https://books.google.co.id/books?id=UVcdEQAAQBAJ

Nurfirdaus, R. E., Barata, M. A., & Prastya, I. W. (2026). *Comparison of SVM and Random Forest for TikTok E10 Fuel Sentiment Analysis*. *10*(2), 1426–1437.

Padmaja, N., Subramaniam, R., & Mohapatra, S. (2024). *Big Data Analytics for the Prediction of Tourist Preferences Worldwide*. Emerald Publishing Limited. https://books.google.co.id/books?id=Lwf0EAAAQBAJ

Phasinam, K., Singh, A. K., Singh, T., & Sharma, M. K. (2022). *Fundamental of Machine Learning*. Blue Rose Publishers. https://books.google.co.id/books?id=-XCJEAAAQBAJ

Pozzi, F. A., Fersini, E., Messina, E., & Liu, B. (2017). *Chapter 1 - Challenges of Sentiment Analysis in Social Networks: An Overview* (F. A. Pozzi, E. Fersini, E. Messina, & B. B. T.-S. A. in S. N. Liu (eds.); pp. 1–11). Morgan Kaufmann. https://doi.org/https://doi.org/10.1016/B978-0-12-804412-4.00001-2

Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics*, *3*(2), 143–157. https://doi.org/https://doi.org/10.1016/j.joi.2009.01.003

Pradana, M., Rubiyanti, N., & Marimon, F. (2024). *Measuring Indonesian young consumers' halal purchase intention of foreign-branded food products*. *2024*. https://doi.org/10.1057/s41599-023-02559-0

Prawiro, A., & Fathudin, F. (2023). Challenges in the Halal Industry Ecosystem: Analyzing the Halal Certification Process for Micro, Small, and Medium Enterprises in Lombok, West Nusa Tenggara. *Mazahib*, *22*, 431–484. https://doi.org/10.21093/mj.v22i2.7010

Prianto, C., & Bunyamin, S. (2020). *Pembuatan aplikasi clustering gangguan jaringan menggunakan metode K-Means clustering*. Kreatif. https://books.google.co.id/books?id=y8TgDwAAQBAJ

Quan-Haase, A., & Sloan, L. (2022). *The SAGE Handbook of Social Media Research Methods*. SAGE Publications. https://books.google.co.id/books?id=Q0mGEAAAQBAJ

Rafika, A., Putri, H., & Wakhidah, N. (2025). *Optimization of K-Means Clusteting with Elbow Method for Identification of TB Prone in Central Java*. *6*(1), 730–736. https://doi.org/10.30596/jcositte.v6i1.21669

Ragavan, A. J. (2025). *Applied Cluster Analysis - Part I: Number of Clusters, Cluster Types, and Quality Evaluation*. Lulu.com. https://books.google.co.id/books?id=e95sEQAAQBAJ

Rianto, Mutiara, A. B., Wibowo, E. P., & Santosa, P. I. (2021). Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation. *Journal of Big Data*, *8*(1), 26. https://doi.org/10.1186/s40537-021-00413-1

Romadhona, W., & Isnain, A. R. (2024). *Analisis sentimen pengguna media sosial terhadap kebijakan kenaikan pajak hiburan menggunakan metode svm (support vector machine)*. *9*(4), 2185–2195.

Salloum, S., Alhumaid, K., Salloum, A., & Shaalan, K. (2024). K-means Clustering of Tweet Emotions : A 2D PCA Visualization Approach. *Procedia Computer Science*, *244*, 30–36. https://doi.org/10.1016/j.procs.2024.10.175

Sandi, F. (2026, April 24). Wajib Halal Berlaku Oktober 2026, Pengusaha: Jangan Mahal & Dipersulit. *CNBC Indonesia*. https://www.cnbcindonesia.com/news/20260424110349-4-729530/wajib-halal-berlaku-oktober-2026-pengusaha-jangan-mahal-dipersulit

Sari, M. D., Saputro, A. R., Anas, M., Kunsah, B., Futuwwah, I., Gigi, F. K., Teknik, F., Kedokteran, F., Kesehatan, F. I., Mikro, U., Engagement, C., Halal, J. P., Lamongan, K., & Mikro, U. (2025). *PEMBERDAYAAN UMKM MELALUI PELATIHAN PENDAMPING PROSES PRODUKSI HALAL PADA PELAKU USAHA KABUPATAEN LAMONGAN*. *4*(2), 186–195. https://doi.org/10.29303/jpimi.v4i2.7094

Setiawan, B. (2024). *A Review of Sentiment Analysis Applications in Indonesia Between 2023-2024*. *08*, 71–83.

Setik, R., Ahmad, R. M. T. R. L., & Marjudi, S. (2021). Exploring Classification For Sentiment Analysis From Halal Based Tweets. *2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, 1–6. https://doi.org/10.1109/AiDAS53897.2021.9574255

Setyowati, R. D., Cahyani, P. G., Annisa, N., & Ni, H. (2024). *Pendampingan Sertifikasi Halal Produk Makanan Usaha Mikro dan Kecil Melalui Website SIHALAL*. *4*, 1–10. https://doi.org/10.35719/ngarsa.v4i1.424

Sholikhah, A., Alrasyid, H., Anwar, S. A., & Ekonomi, F. (2025). *Analisis Dampak Literasi*

Halal dan Biaya Sertifikasi Halal Terhadap Minat Melakukan Sertifikasi Halal Pada Sektor UMKM Kuliner. *8*(1), 304–313.

Simbolon, S. E. A., & Hidayat, N. W. (2021). Prosedur dan Problematika Sertifikasi Halal Di Indonesia. *Masyrif: Jurnal Ekonomi, Bisnis Dan Manajemen*, *2*(1), 118–132.

Simmonds, D., & Chowdhury, A. (2022). *K-means Clustering of Sentiment Analysis within Online Reviews -An NLP Approach*.

Tan, Y., Shi, Y., & Coello, C. C. (2014). *Advances in Swarm Intelligence: 5th International Conference, ICSI 2014, Hefei, China, October 17-20, 2014, Proceedings, Part I*. Springer International Publishing. https://books.google.co.id/books?id=CMxsBAAAQBAJ

Tarigan, W. (2026). *Implementation System for Cluster Analysis and Data Sentiment Using the K-Means Method to Determine the Most Discussed Topics*. *12*(3), 584–592. https://doi.org/10.29303/jppipa.v12i3.14838

Wang, W. (2024). *Principles of Machine Learning: The Three Perspectives*. Springer Nature Singapore. https://books.google.co.id/books?id=3JosEQAAQBAJ

Warto, A. ., Waluya, A., & Rachman, A. (2025). *Analisis Kesiapan UMKM dalam Menghadapi Sertifikasi Halal Wajib 2026 di Indonesia*. *1*, 72–92.

Wicaksana, R. S., Heksaputra, D., S, T. A., & Febri, F. (2023). *Pendekatan K-Means Clustering Metode Elbow Pada Analisis Motivasi Pengunjung Festival Halal JHF # 2*. *9*(03), 4162–4176.

Widodo, A. C., & Nuryanto, Y. (2024). *Pengawasan Jaminan Produk Halal (JPH) Oleh Badan Penyelenggara Jaminan Produk Halal (BPJPH)*. *7*(10), 12368–12377.

Wiryono, S. (2024). *Kewajiban Sertifikasi Halal Diberlakukan Secara Resmi Mulai 18 Oktober*. https://nasional.kompas.com/read/2024/10/18/18501051/kewajiban-sertifikasi-halal-diberlakukan-secara-resmi-mulai-18-oktober

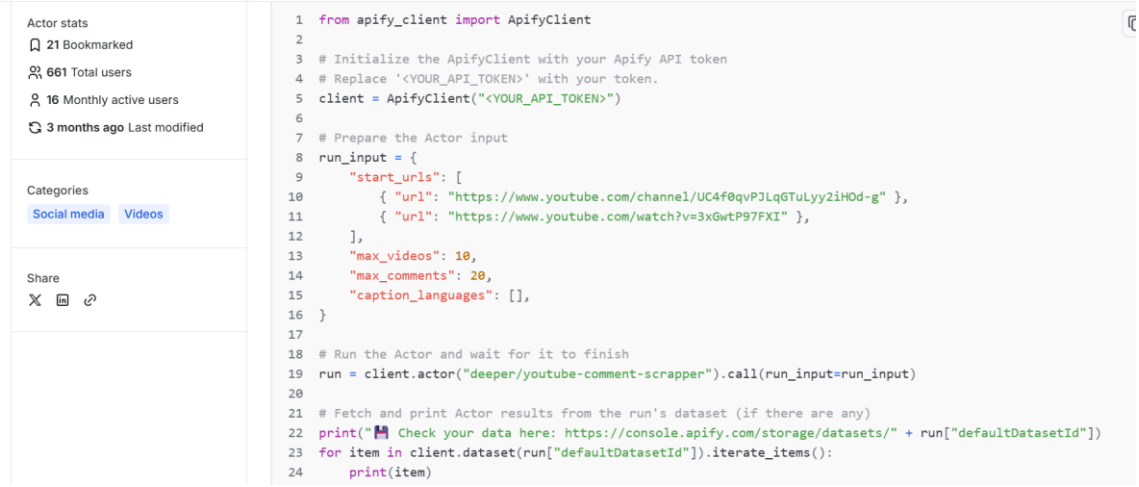
Yam, J. H. (2024). *JURNAL EMPIRE Kajian Penelitian : Tinjauan Literatur Sebagai Metode Penelitian*. *4*(1), 61–71.

Yessy Asri, S. T. M., & Dr. Dra. Dwina Kuswardani, M. K. (2024). *MACHINE LEARNING & DEEP LEARNING: Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi*. Uwais Inspirasi indonesia . https://books.google.co.id/books?id=Yu7uEAAAQBAJ

Zannettou, S., Nemes-nemeth, O., Goetzen, A., Gummadi, K. P., & Redmiles, E. M. (2024). Analyzing User Engagement with TikTok ' s Short Format Video Recommendations using Data Donations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Vol. 1, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642433

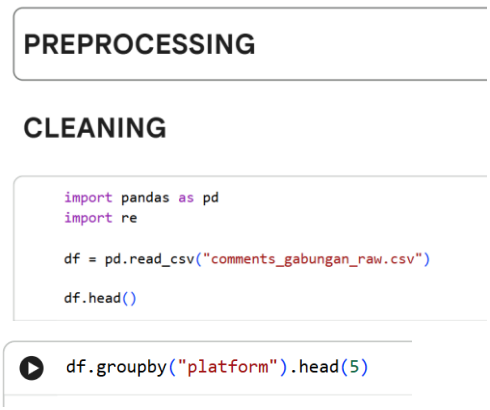
LAMPIRAN

A. Pengumpulan Data




```python
from apify_client import ApifyClient

# Initialize the ApifyClient with your Apify API token
# Replace '<YOUR_API_TOKEN>' with your token.
client = ApifyClient("<YOUR_API_TOKEN>")

# Prepare the Actor input
run_input = { "postUrls": ["https://www.instagram.com/p/DNEYQDvAaYh/?img_index=1"] }

# Run the Actor and wait for it to finish
run = client.actor("rakser/instagram-comments-scrapper").call(run_input=run_input)

# Fetch and print Actor results from the run's dataset (if there are any)
print("💾 Check your data here: https://console.apify.com/storage/datasets/" + run["defaultDatasetId"])
for item in client.dataset(run["defaultDatasetId"]).iterate_items():
    print(item)
```

B. Kode *Pre-processing* Data



```python
def clean_text(text):
    if pd.isna(text):
        return ""

    text = str(text)

    # hapus URL
    text = re.sub(r'http\S+|www\S+|https\S+', ' ', text)

    # hapus mention username
    text = re.sub(r'@\w+', ' ', text)

    # hapus hashtag, tapi teksnya tetap disimpan
    text = re.sub(r'#', '', text)

    # hapus emoji dan karakter non-ASCII
    text = text.encode('ascii', 'ignore').decode('ascii')

    # hapus angka
    text = re.sub(r'\d+', ' ', text)

    # hapus tanda baca dan simbol
    text = re.sub(r'[^a-zA-Z\s]', ' ', text)

    # hapus spasi berlebih
    text = re.sub(r'\s+', ' ', text).strip()

    return text
```

```python
df["komentar_cleaning"] = df["komentar"].apply(clean_text)

df = df[df["komentar_cleaning"] != ""]

# hapus data duplikat berdasarkan hasil cleaning
df = df.drop_duplicates(subset=["komentar_cleaning"])

# reset index
df = df.reset_index(drop=True)

df[["komentar", "komentar_cleaning"]].head(10)
```

```python
df.to_csv("comments_cleaning.csv", index=False)

print("Jumlah data setelah cleaning:", len(df))
```

```
Jumlah data setelah cleaning: 4846
```

```python
from google.colab import files
files.download("comments_cleaning.csv")
```

CASE FOLDING

```python
import pandas as pd
df = pd.read_csv("comments_cleaning.csv")

df["komentar_casefold"] = df["komentar_cleaning"].str.lower()

df[["komentar_cleaning", "komentar_casefold"]].head(10)
```

NORMALISASI SLANG

```python
import pandas as pd

df = pd.read_csv("comments_casefold.csv")

slang_dict = {
    "gak": "tidak", "ga": "tidak", "nggak": "tidak", "gk": "tidak",
    "tdk": "tidak", "tak": "tidak", "ngak": "tidak", "gax": "tidak",

    "yg": "yang", "yng": "yang",
    "dr": "dari",
    "krn": "karena",
    "karna": "karena",

    "bgt": "banget", "bgtt": "banget", "bngt": "banget",
    "bangetttt": "banget",

    "aja": "saja", "ajha": "saja", "aj": "saja",
    "doang": "saja",

    "kalo": "kalau", "kl": "kalau",
    "trs": "terus", "trus": "terus", "teros": "terus", "terusssss": "terus",

    "dgn": "dengan",
    "sm": "sama", "sma": "sama",

    "udh": "sudah", "sdh": "sudah",
    "blm": "belum", "belm": "belum", "blom": "belum", "blum": "belum",

    "org": "orang",
    "gw": "saya", "gue": "saya", "gua": "saya", "sy": "saya",
    "lu": "kamu", "loe": "kamu", "lo": "kamu",

    "emg": "memang",
    "knp": "kenapa",

    "ni": "nih", "donk": "dong", "deh": "deh", "lah": "lah",
    "dh": "sudah", "sdh": "sudah",
    "dpt": "dapat", "dapet": "dapat",
    "dll": "dan lainnya",
    "syethan": "setan", "setannnnn": "setan",

    "gratisssss": "gratis",
    "cepet": "cepat", "cpt": "cepat",
    "tu": "itu", "ituuuuuuuu": "itu",
    "rakyatsttt": "rakyat", "rkyat": "rakyat",

    "kalo": "kalau", "klo": "kalau", "klw": "kalau", "kl": "kalau", "klau": "kalau",
    "pdhl": "padahal",
```

TOKENIZATION

```
import pandas as pd

df = pd.read_csv("comments_normalisasi.csv")

def tokenize_text(text):
    if pd.isna(text):
        return []
    return str(text).split()

df["komentar_token"] = df["komentar_normalisasi"].apply(tokenize_text)

df[["komentar_normalisasi", "komentar_token"]].head(10)
```

	komentar_normalisasi	komentar_token
0	daftar nya dimana	[daftar, nya, dimana]
1	apakah pemerintah tidak bisa mencari penghasil...	[apakah, pemerintah, tidak, bisa, mencari, pen...
2	telat udah mereka lebih senang mengadu domba d...	[telat, udah, mereka, lebih, senang, mengadu, ...
3	menghalalkan semua agama	[menghalalkan, semua, agama]
4	tidak cuma gula tapi garam lemak dan lainnya h...	[tidak, cuma, gula, tapi, garam, lemak, dan, l...
5	wah belum cair cair nih keterlaluan	[wah, belum, cair, cair, nih, keterlaluan]
6	supaya konsumen jelas pakai bahasa indonesia s...	[supaya, konsumen, jelas, pakai, bahasa, indon...
7	babi akan di halalkan kalau bayar sertifikasi ...	[babi, akan, di, halalkan, kalau, bayar, serti...
8	waktunya anies presiden	[waktunya, anies, presiden]
9	oh intinya sertifikat nya halal	[oh, intinya, sertifikat, nya, halal]

```
df.to_csv("comments_tokenization.csv", index=False)

print("Tokenization selesai!")
print("Jumlah data:", len(df))
```

```
Tokenization selesai!
Jumlah data: 4046
```

```
from google.colab import files
files.download("comments_tokenization.csv")
```

```
!pip install Sastrawi
```

```
Collecting Sastrawi
  Downloading Sastrawi-1.0.1-py2.py3-none-any.whl.metadata (909 bytes)
Downloading Sastrawi-1.0.1-py2.py3-none-any.whl (209 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 209.7/209.7 kB 4.6 MB/s eta 0:00:00
Installing collected packages: Sastrawi
Successfully installed Sastrawi-1.0.1
```

```
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

factory = StopWordRemoverFactory()
stopwords = set(factory.get_stop_words())

custom_stopwords = {
    "nya", "sih", "dong", "nih", "deh", "lah",
    "aja", "ya", "kan", "kak", "bang", "bro",
    "guys", "wkwk", "haha", "hehe"
}

stopwords.update(custom_stopwords)

# kata penting untuk sentimen dan konteks kebijakan jangan dihapus
important_words = {"tidak", "bukan", "jangan", "harus", "wajib", "boleh", "kentut", "dapat", "supaya", "sudah", "dimana"}
stopwords = stopwords - important_words

def remove_stopwords(tokens):
    return [word for word in tokens if word not in stopwords]
```

```
penting = ["tidak", "bukan", "jangan", "harus", "wajib", "boleh", "sudah", "supaya", "kentut", "dapat", "dimana"]

for kata in penting:
    stopwords.discard(kata)
```

STEMMING

```
!pip install Sastrawi
```

```
Requirement already satisfied: Sastrawi in /usr/local/lib/python3.12/dist-packages (1.0.1)
```

```
import pandas as pd
import ast
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory


df = pd.read_csv("comments_stopword.csv")


df["komentar_stopword"] = df["komentar_stopword"].apply(ast.literal_eval)


factory = StemmerFactory()
stemmer = factory.create_stemmer()


def stemming_tokens(tokens):
    return [stemmer.stem(word) for word in tokens]


df["komentar_stemming"] = df["komentar_stopword"].apply(stemming_tokens)


df["komentar_final"] = df["komentar_stemming"].apply(lambda x: " ".join(x))


df[["komentar_stopword", "komentar_stemming", "komentar_final"]].head(10)
```

```python
import pandas as pd

log_preprocessing = []

# Load data from the respective CSVs to get the lengths at each stage
df_raw_loaded = pd.read_csv("comments_gabungan_raw.csv")
log_preprocessing.append(["Data awal", len(df_raw_loaded)])

df_clean_loaded = pd.read_csv("comments_cleaning.csv")
log_preprocessing.append(["Setelah cleaning", len(df_clean_loaded)])

df_stopword_loaded = pd.read_csv("comments_stopword.csv")
log_preprocessing.append(["Setelah stopword removal", len(df_stopword_loaded)])

# 'df' at this point contains the data after stemming
log_preprocessing.append(["Setelah stemming", len(df)])

log_df = pd.DataFrame(log_preprocessing, columns=["Tahap", "Jumlah Data"])
log_df
```

	Tahap	Jumlah Data
0	Data awal	4908
1	Setelah cleaning	4846
2	Setelah stopword removal	4846
3	Setelah stemming	4841

C. Kode TF-IDF

```python
from sklearn.feature_extraction.text import TfidfVectorizer
import pandas as pd

df = pd.read_csv("comments_stemming.csv")

vectorizer = TfidfVectorizer(
    max_features=5000,   # batas jumlah fitur (boleh disesuaikan)
    min_df=2,            # kata muncul minimal 2 kali
    max_df=0.8           # buang kata terlalu sering
)

X = vectorizer.fit_transform(df["komentar_final"])

print("Shape TF-IDF:", X.shape)
```

```
Shape TF-IDF: (4841, 2803)
```

```python
# ubah ke dataframe biar kebaca
tfidf_df = pd.DataFrame(
    X.toarray(),
    columns=vectorizer.get_feature_names_out()
)

tfidf_df.head()
```

	aa	aamiin	abah	abang	abis	abu	acc	ad	ada	adab	...	yes	yesus	yo	yoi	you	youtube	yuk	yup	zaman	zina
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows × 2803 columns

D. Kode Penentuan Jumlah Cluster

```python
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score, davies_bouldin_score


scores = []

for k in range(2, 11):
    kmeans = KMeans(
        n_clusters=k,
        random_state=42,
        n_init=10
    )

    labels = kmeans.fit_predict(X)

    scores.append({
        "k": k,
        "inertia_wcss": kmeans.inertia_,
        "silhouette_score": silhouette_score(X, labels),
        "davies_bouldin_index": davies_bouldin_score(X.toarray(), labels)
    })

scores_df = pd.DataFrame(scores)

display(scores_df)


plt.figure(figsize=(8, 5))
plt.plot(scores_df["k"], scores_df["inertia_wcss"], marker="o")
plt.title("Elbow Method")
plt.xlabel("Jumlah Cluster (k)")
plt.ylabel("Inertia / WCSS")
plt.xticks(scores_df["k"])
plt.grid(True)
plt.show()
```

E. Pengelompokkan dengan *K-Means Clustering*

```
from sklearn.cluster import KMeans

k_optimal = 6

kmeans_final = KMeans(
    n_clusters=k_optimal,
    random_state=42,
    n_init=10
)

df["cluster"] = kmeans_final.fit_predict(X)

# Cek hasil awal
df[["komentar_final", "cluster"]].head()
```

	komentar_final	cluster
0	daftar	2
1	pemerintah tidak nyari hasil sendiri majakin o...	2
2	telat udah lbh semang adu domba drpd berita be...	2
3	halal semua agama	3
4	tidak cuma gula garam lemak lain harus	2

```
cluster_counts = df["cluster"].value_counts().sort_index()

print("Jumlah data pada setiap cluster:")
print(cluster_counts)
```

```
Jumlah data pada setiap cluster:
cluster
0     550
1      85
2    2836
3     801
4     192
5     377
Name: count, dtype: int64
```

```
cluster_summary = cluster_counts.reset_index()
cluster_summary.columns = ["cluster", "jumlah_komentar"]
cluster_summary
```

F. Kode Kata Dominan

MELIHAT KATA DOMINAN

```
import numpy as np
import pandas as pd

terms = vectorizer.get_feature_names_out()

top_terms_per_cluster = []

for cluster_id in sorted(df["cluster"].unique()):
    cluster_indices = df[df["cluster"] == cluster_id].index

    # Ambil rata-rata bobot TF-IDF pada cluster tersebut
    cluster_tfidf = X[cluster_indices].mean(axis=0)
    cluster_tfidf = np.asarray(cluster_tfidf).flatten()

    # Ambil 15 kata dengan bobot tertinggi
    top_indices = cluster_tfidf.argsort()[::-1][:15]
    top_terms = [terms[i] for i in top_indices]
    top_scores = [cluster_tfidf[i] for i in top_indices]

    top_terms_per_cluster.append({
        "cluster": cluster_id,
        "kata_dominan": ", ".join(top_terms),
        "bobot_tfidf": top_scores
    })

top_terms_df = pd.DataFrame(top_terms_per_cluster)
top_terms_df[["cluster", "kata_dominan"]]
```

	cluster	kata_dominan
0	0	sertifikat, gratis, halal, min, kalau, bayar, ...
1	1	cuan, ladang, baru, tambah, cari, sertifikasi,...
2	2	tidak, negara, kalau, jadi, buat, indonesia, h...
3	3	halal, tidak, non, makan, kalau, label, harus,...
4	4	duit, ujung, uud, nyari, ladang, sertifikasi, ...
5	5	babi, jual, gimana, kalau, halal, daging, maka...

G. Kode Profilisasi Cluster

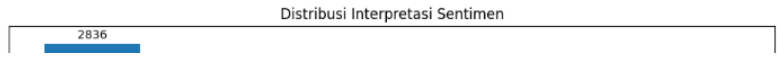
PROFILISASI CLUSTER

```
cluster_profile = cluster_summary.merge(
    top_terms_df[["cluster", "kata_dominan"]],
    on="cluster",
    how="left"
)

cluster_profile
```

	cluster	jumlah_komentar	kata_dominan
0	0	550	sertifikat, gratis, halal, min, kalau, bayar, ...
1	1	85	cuan, ladang, baru, tambah, cari, sertifikasi,...
2	2	2836	tidak, negara, kalau, jadi, buat, indonesia, h...
3	3	801	halal, tidak, non, makan, kalau, label, harus,...
4	4	192	duit, ujung, uud, nyari, ladang, sertifikasi, ...
5	5	377	babi, jual, gimana, kalau, halal, daging, maka...

```
# ==============================
# PROFILISASI CLUSTER DARI FILE EXCEL
# ==============================

import pandas as pd

# Baca file Excel profilisasi cluster
profilisasi_cluster = pd.read_excel("Berdasarkan hasil setiap cluster berikut, buatkan....xlsx")

# Samakan nama kolom dari file Excel
profilisasi_cluster = profilisasi_cluster.rename(columns={
    "Cluster": "cluster",
    "Label Nama Cluster": "labeling",
    "Interpretasi Sentimen": "interpretasi_sentimen",
    "Deskripsi Karakteristik Cluster": "deskripsi_cluster"
})

# Pastikan kolom cluster bertipe integer
cluster_profile["cluster"] = cluster_profile["cluster"].astype(int)
profilisasi_cluster["cluster"] = profilisasi_cluster["cluster"].astype(int)

# ==============================
# GABUNGKAN JUMLAH KOMENTAR, KATA DOMINAN, DAN PROFILISASI
# ==============================

cluster_profile = cluster_profile.merge(
    profilisasi_cluster,
```

```
# Samakan nama kolom dari file Excel
profilisasi_cluster = profilisasi_cluster.rename(columns={
    "Cluster": "cluster",
    "Label Nama Cluster": "labeling",
    "Interpretasi Sentimen": "interpretasi_sentimen",
    "Deskripsi Karakteristik Cluster": "deskripsi_cluster"
})

# Pastikan kolom cluster bertipe integer
cluster_profile["cluster"] = cluster_profile["cluster"].astype(int)
profilisasi_cluster["cluster"] = profilisasi_cluster["cluster"].astype(int)

# ==============================
# GABUNGKAN JUMLAH KOMENTAR, KATA DOMINAN, DAN PROFILISASI
# ==============================

cluster_profile = cluster_profile.merge(
    profilisasi_cluster,
    on="cluster",
    how="left"
)

# ==============================
# SUSUN ULANG KOLOM AGAR RAPI
# ==============================

cluster_profile = cluster_profile[
    [
        "cluster",
        "jumlah_komentar",
        "kata_dominan",
        "labeling",
        "interpretasi_sentimen",
        "deskripsi_cluster"
    ]
]

# Tampilkan hasil profilisasi cluster
cluster_profile
```

H. Kode Distribusi Sentimen

```
import matplotlib.pyplot as plt
from google.colab import files

plt.figure(figsize=(10, 6))

plt.bar(sentimen_counts.index, sentimen_counts.values)

plt.title("Distribusi Interpretasi Sentimen")
plt.xlabel("Interpretasi Sentimen")
plt.ylabel("Jumlah Komentar")

plt.xticks(rotation=30, ha="right")

for i, v in enumerate(sentimen_counts.values):
    plt.text(i, v + 20, str(v), ha="center", va="bottom")

plt.tight_layout()

# Simpan gambar SEBELUM plt.show()
plt.savefig("grafik_distribusi_sentimen.png", dpi=300, bbox_inches="tight")

plt.show()

# Download gambar
files.download("grafik_distribusi_sentimen.png")
```