

**DESAIN USULAN PROGRAM PERCEPATAN SERTIFIKASI
HALAL INDUSTRI DI INDONESIA BERBASIS ANALISIS
SENTIMEN KONSUMEN DI MEDIA SOSIAL TIKTOK**

TUGAS AKHIR

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



Nama : Qonita Silvia Faizun
No. Mahasiswa : 22522110

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

2026

PERNYATAAN KEASLIAN

PERNYATAAN KEASLIAN

Saya mengakui bahwa tugas akhir ini adalah hasil karya saya sendiri kecuali kutipan dan ringkasan yang seluruhnya sudah saya jelaskan sumbernya. Jika dikemudian 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



(Qonita Silvia Faizun)
22522110

SURAT BUKTI PENELITIAN



FAKULTAS
TEKNOLOGI INDUSTRI

Gedung K1, Jalan Marmar
Kampus Terpadu Universitas Islam Indonesia
Jl. Kaliurang km 14,5 Yogyakarta 55584
T. (0274) 898444 ext. 4100, 4101
F. (0274) 899087
E. lib@uii.ac.id
W. lib.uii.ac.id

SURAT KETERANGAN PENELITIAN

Nomor: 015/Ka.Lab.Datamin/70/Lab.Datamin/V/2026

Assalamu'alaikum Warahmatullahi Wabarakatuh

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

Nama : Qonita Silvia Faizun
No. Mhs : 22522110
Dosen Pembimbing : Dr. Harwati, S.T., M.T.

Telah selesai melaksanakan penelitian yang berjudul " Desain Usulan Program Percepatan Sertifikasi Halal Industri di Indonesia berbasis Analisis Sentimen Konsumen di Media Sosial TikTok" di Laboratorium Data Mining, Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 2 Januari 2026 sampai dengan tanggal 7 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

**DESAIN USULAN PROGRAM PERCEPATAN SERTIFIKASI HALAL
INDUSTRI DI INDONESIA BERBASIS ANALISIS SENTIMEN KONSUMEN DI
MEDIA SOSIAL TIKTOK**



TUGAS AKHIR

Disusun Oleh:

Nama : Qonita Silvia Faizun

No. Mahasiswa : 22522110

Yogyakarta, 13 Mei 2026

Dosen Pembimbing

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

LEMBAR PENGESAHAN DOSEN PENGUJI

DESAIN USULAN PROGRAM PERCEPATAN SERTIFIKASI HALAL INDUSTRI DI INDONESIA BERBASIS ANALISIS SENTIMEN KONSUMEN DI MEDIA SOSIAL TIKTOK

TUGAS AKHIR

Disusun Oleh :

Nama : Qonita Silvia Faizun

No. Mahasiswa : 22522110

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, 17 Juni 2026

Tim Penguji

Dr. Harwati, S.T, M.T.

Ketua

Dr. Joko Sulistio, S.T., M.Sc., M.T

Anggota I

Ir. Muchamad Sugarindra, S.T., M.T.I., IPM.

Anggota II

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

Bismillahirrahmanirrahim

Dengan rasa syukur yang tulus kepada Allah SWT atas nikmat, bimbingan, dan kekuatannya yang tidak ada habisnya selama penyelesaian Tugas Akhir ini, saya persembahkan karya ini kepada orang tua dan keluarga tercinta atas doa, kasih sayang, dan dukungan mereka yang tidak pernah padam, serta kepada para dosen, teman, dan semua orang yang telah menyemangati dan menemani saya selama perjalanan ini. Semoga penelitian ini membawa berkah dan manfaat untuk langkah yang lebih bermakna menuju masa depan yang lebih baik.

MOTTO

“Allah tempat meminta segala sesuatu”

(QS. Al -ikhlas: 2)

“Dia (Musa) menjawab, ‘Sekali-kali akan tersusul; sesungguhnya Tuhanku bersamaku, kelak dia akan memberi petunjuk kepadaku”

(QS. Asy-Syu’ara: 62)

“Kesuksesan adalah hasil dari usaha-usaha kecil yang diulang setiap hari”

(Robert Collier)

KATA PENGANTAR

Bismillahirrahmanirrahim,

Alhamdulillah rabbil'alam, segala puji dan syukur hanya kepada Allah SWT atas nikmat, petunjuk, dan rahmat-Nya yang telah memungkinkan penulis untuk menyelesaikan proyek Tugas Akhir yang berjudul “Desain Usulan Program Strategi Percepatan Sertifikasi Halal di Indonesia berbasis Analisis Sentimen Konsumen di Media Sosial TikTok” dengan sukses sebagai persyaratan memperoleh gelar sarjana. Penulis juga menyadari bahwa Tugas Akhir ini tidak mungkin dapat diselesaikan tanpa dukungan, bantuan, doa, dan dorongan dari banyak orang.

Oleh karena itu, penulis ingin menyampaikan penghargaan yang tulus kepada dosen, pembimbing, keluarga, teman, dan semua orang yang telah berkontribusi dan memberikan motivasi sepanjang perjalanan ini. Dengan ini, penulis ingin berterima kasih 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 Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia.
3. Bapak Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM. Selaku Ketua Program Studi Teknik Industri Program Sarjana Fakultas Teknologi Industri, Universitas Islam Indonesia.
4. Ibu Dr. Harwati, S.T, M.T. selaku dosen pembimbing tugas akhir yang telah memberikan ilmu, bimbingan, dan arahan dengan tulus sehingga penulis dapat menyelesaikan tugas akhir ini dengan selesai.
5. Kedua Orang Tua penulis, Bapak Faizun dan Ibu Titi Subarkah, yang selalu memberikan dukungan baik materil maupun doa yang tidak ada habisnya kepada penulis.
6. Teman-teman seperjuangan penulis yaitu teman terdekat dari Teknik Industri angkatan 2022 yang telah memberikan dukungan dan semangat dalam melaksanakan tugas akhir ini.
7. Semua pihak yang telah membantu penulisann tugas akhir ini yang tidak dapat disebutkan satu per satu.

Penulis masih menyadari jika penulisan laporan tugas akhir ini masih terdapat adanya kekurangan dan jauh dari kata sempurna. Oleh karena itu, penulis berharap kritik dan saran yang berguna dengan tujuan untuk memperbaiki tugas akhir ini. Terakhir, harapan penulis semoga laporan tugas akhir ini bermanfaat bagi penulis dan pembaca.

Yogyakarta, 13 Mei 2026



Qonita Silvia Faizun

NIM 22522110

ABSTRAK

Kebijakan sertifikasi halal wajib di Indonesia bertujuan untuk memastikan jaminan produk halal dan meningkatkan kepercayaan konsumen, terutama di kalangan konsumen Muslim. Namun, jumlah UMKM yang memiliki sertifikasi halal masih relatif rendah dibandingkan dengan jumlah total UMKM di Indonesia. Opini publik mengenai kebijakan ini banyak diutarakan di Media Sosial TikTok. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen konsumen terhadap kebijakan sertifikasi halal berdasarkan komentar TikTok dan mengembangkan rekomendasi atau desain usulan untuk mempercepat program sertifikasi halal.

Penelitian ini menerapkan metode *Text Mining* dengan tahap pre-processing, termasuk case folding, cleaning, tokenisasi, stopword removing, dan stemming. Data dianalisis menggunakan metode pembobotan TF-IDF, K-Means *Clustering*, Metode Elbow, dan *Lexicon-based* analysis. Sebanyak 1.100 komentar dikumpulkan melalui proses *scraping* dari konten yang berkaitan dengan kebijakan sertifikasi halal.

Hasil menunjukkan bahwa pengelompokan optimal membentuk lima kelompok dengan karakteristik sentimen yang berbeda. Sentimen positif terutama mencerminkan dukungan terhadap sertifikasi halal dan manfaatnya, sementara sentimen negatif berkaitan dengan biaya sertifikasi, kompleksitas proses, dan keraguan tentang dampaknya terhadap penjualan. Berdasarkan penelitian ini, studi ini mengusulkan strategi untuk mempercepat sertifikasi halal melalui sosialisasi digital, pendidikan publik, dan program bantuan untuk UMKM.

Kata Kunci: Sertifikasi Halal, Analisis Sentimen, TikTok, *Clustering K-Means*, *Lexicon-based Analysis*

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	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
BAB I	
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	6
1.4 Manfaat Penelitian	6
1.5 Batasan Penelitian	6
1.6 Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA	9
2.1 Kajian Literatur	9
2.1.1 Posisi Penelitian	12
2.2 Landasan Teori	16
2.2.1 Kebijakan Sertifikasi Halal	16
2.2.2 TikTok	17
2.2.3 Text Mining	18
2.2.4 Analisis Sentimen	20
2.2.5 Pembobotan TF-IDF	20
2.2.6 Clustering	21
2.2.7 Algoritma K-Means	22
2.2.8 Metode Elbow	23
2.2.9 Lexicon-based Analysis	23
2.2.10 WordCloud	25
BAB III METODE PENELITIAN	26
3.1 Objek dan Subjek Penelitian	26
3.2 Pengumpulan Data Penelitian	26
3.3 Metode Pengumpulan Data Penelitian	26
3.4 Jenis dan Sumber Data	27
3.5 Variabel Penelitian	27
3.6 Metode Analisis Data	27
3.7 Alur Penelitian	29
BAB IV PENGUMPULAN DAN PENGOLAHAN DATA	35
4.1 Pengumpulan Data	35
4.2 Pengolahan Data	39
4.2.1 Pre-Processing Data	39
4.3 Identifikasi Ciri Khas setiap <i>Cluster</i>	45
4.3.1 Cluster 1	45

4.3.2	Cluster 2.....	49
4.3.3	Cluster 3.....	52
4.3.4	Cluster 4.....	56
4.3.5	Cluster 5.....	59
4.4	Hasil Ringkasan Ciri Khas Sentimen tiap-tiap <i>Cluster</i>	63
BAB V PEMBAHASAN.....		67
5.1	Profiling <i>Cluster</i>	67
5.1.1	Cluster 1.....	67
5.1.2	Cluster 2.....	67
5.1.3	Cluster 3.....	67
5.1.4	Cluster 4.....	68
5.1.5	Cluster 5.....	68
5.2	Desain Usulan Strategi Percepatan Sertifikasi Halal	69
5.2.1	Strategi 1 Sosialisasi Transparasi Biaya	71
5.2.2	Strategi 2 Kampanye Testimoni UMKM	73
5.2.3	Strategi 3 Kolaborasi pemerintah dengan influencer.....	75
5.2.4	Strategi 4 Kampanye “Kenali Bahan non-Halal”	77
5.3	Limitasi Penelitian	78
BAB VI PENUTUP.....		80
6.1	Kesimpulan	80
6.2	Saran	81
DAFTAR PUSTAKA		82
LAMPIRAN		A-1

DAFTAR TABEL

Tabel 2. 1 Posisi Penelitian.....	12
Tabel 4. 1 <i>Case Folding</i> (sebelum dan Sesudah)	40
Tabel 4. 2 <i>Tokenizing</i>	41
Tabel 4. 3 <i>Stopword Removal</i>	41
Tabel 4. 4 <i>Stemming</i>	42
Tabel 4. 5 Frekuensi dan Pembobotan Kata	46
Tabel 4. 6 Frekuensi dan Pembobotan Kata Gabungan	48
Tabel 4. 7 Frekuensi dan Pembobotan Kata	50
Tabel 4. 8 Frekuensi dan Pembobotan Kata Gabungan	51
Tabel 4. 9 Frekuensi dan Pembobotan Kata	53
Tabel 4. 10 Frekuensi dan Pembobotan Kata Gabungan	55
Tabel 4. 11 Frekuensi dan Pembobotan Kata	57
Tabel 4. 12 Frekuensi dan Pembobotan Kata Gabungan	58
Tabel 4. 13 Frekuensi dan Pembobotan Kata	60
Tabel 4. 14 Frekuensi dan Pembobotan Kata Gabungan	62
Tabel 4. 15 Ciri Khas Tiap <i>Cluster</i>	63
Tabel 4. 16 Proporsi Sentimen.....	66
Tabel 5. 1 Strategi Percepatan Sertifikasi Halal	69

DAFTAR GAMBAR

Gambar 1. 1 Data perbandingan total UMKM di Indonesia dan UMKM bersertifikasi halal 2025	2
Gambar 1. 2 Dampak sertifikasi penjualan terhadap omzet penjualan	3
Gambar 2. 1 Proses <i>Text Mining</i>	18
Gambar 2. 2 Contoh <i>WordCloud</i>	25
Gambar 3. 1 Alur Penelitian (1).....	30
Gambar 3. 2 Alur Penelitian (2).....	31
Gambar 4. 1 <i>Interface Platform Apify</i>	35
Gambar 4. 2 <i>Log In</i> di Apify.....	35
Gambar 4. 3 <i>search actors</i>	36
Gambar 4. 4 menuliskan TikTok Comment <i>Scraper</i>	36
Gambar 4. 5 TikTok Comment <i>Scraper</i> actors.....	37
Gambar 4. 6 Proses <i>Running</i>	37
Gambar 4. 7 <i>Running</i> selesai dan <i>Export file</i>	38
Gambar 4. 8 Download File csv	38
Gambar 4. 9 File format csv yang sudah di download	39
Gambar 4. 10 <i>Text Cleaning</i>	40
Gambar 4. 11 Bobot Frekuensi TF-IDF	43
Gambar 4. 12 Jumlah <i>Cluster</i>	43
Gambar 4. 13 Pemeriksaan Data Nan (null).....	44
Gambar 4. 14 Grafik Jumlah Clustering	44
Gambar 4. 15 Jumlah masing-masing <i>cluster</i>	45
Gambar 4. 16 Visualisasi <i>WordCloud Cluster 1</i>	46
Gambar 4. 17 Visualisasi <i>WordCloud Cluster 2</i>	49
Gambar 4. 18 Visualisasi <i>WordCloud Cluster 3</i>	53
Gambar 4. 19 Visualisasi <i>WordCloud Cluster 4</i>	56
Gambar 4. 20 Visualisasi <i>WordCloud Cluster 5</i>	60
Gambar 5. 1 Implementasi Desain Komunikasi terhadap Publik	71
Gambar 5. 2 Contoh Implementasi Program Strategi 1	72
Gambar 5. 3 Contoh Implementasi Program Strategi 1 (2)	73
Gambar 5. 4 Contoh Implementasi Program Strategi 2	74
Gambar 5. 5 Contoh Implementasi Program Strategi 2(2)	75
Gambar 5. 6 Contoh Implementasi Program Strategi 3	76
Gambar 5. 7 Contoh Implementasi Program Strategi 4	78

BAB I

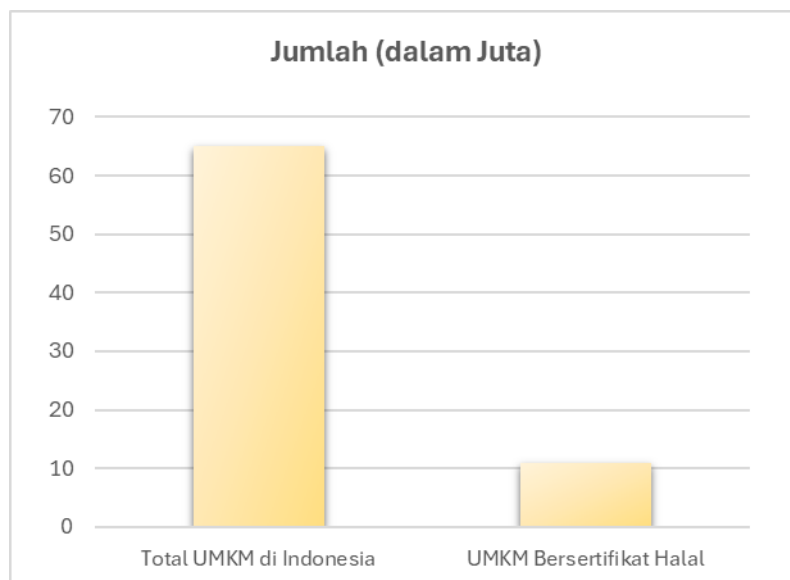
PENDAHULUAN

1.1 Latar Belakang

Produk halal menjadi topik yang mendapat banyak sorotan di beberapa tahun terakhir ini, terutama di negara dengan mayoritas Muslim terbesar seperti Indonesia (Fathoni, 2021). Produk halal identik dengan diperlukannya sertifikasi halal sebagai pedoman umat Muslim saat ini. Sertifikasi halal merupakan hal penting bagi bisnis yang bertujuan untuk memperluas jangkauan pasar dan membangun kepercayaan dengan konsumen yang memprioritaskan kepatuhan agama dalam pembelian mereka (Arif et al., 2023). Pemerintah Indonesia telah menetapkan kebijakan kewajiban sertifikasi halal di Indonesia sejak 2019. Kebijakan kewajiban sertifikasi halal ini diatur dalam Undang-Undang Nomor 33 Tahun 2014 tentang Jaminan Produk Halal (UU JPH) ayat 10, Sertifikasi halal adalah pengakuan kehalalan suatu produk yang dikeluarkan oleh BPJPH (Badan Penyelenggara Produk Halal) berdasarkan fatwa halal tertulis yang dikeluarkan Majelis Ulama Indonesia. Serta, kebijakan ini dimulai sejak 2019 dan dilaksanakan secara bertahap, dengan tahapan pertama berakhir pada 17 Oktober 2024. Serta, menurut Halal MUI saat ini pada Januari 2026, kewajiban ini sudah berlaku penuh untuk usaha menengah dan besar.

Pemerintah memiliki maksud dalam memulainya penerapan kewajiban sertifikasi halal dengan tujuan seperti pada UU Nomor 33 Tahun 2014 tentang Jaminan Produk Halal (UU JPH) ayat 10 ini yang mewajibkan pelaku usaha membuat sertifikasi produk agar memenuhi standar halal Indonesia dan melindungi konsumen Muslim mayoritas. Dari adanya tujuan tersebut, pemerintah Indonesia berhasil menerapkan pertumbuhan kewajiban sertifikasi halal dari tahun 2020 hingga tahun 2025 yang mencapai 10,98 juta sertifikasi oleh Databoks dan BPJPH. Namun, berdasarkan total data UMKM yang ada di Indonesia, penerapan sertifikat halal ini masih belum tercapai secara maksimal, hal ini ditunjukkan oleh frekuensi jumlah UMKM di Indonesia yang mencapai 65 juta unit usaha menurut Badan Pusat Statistik (BPS) pada tahun 2025. Untuk itu, pertumbuhan produk/UMKM bersertifikat halal di Indonesia dari tahun ke tahun selama tahun 2020 hingga tahun 2025 memiliki pertumbuhan yang meningkat tetapi belum terlaksana secara merata dilihat dari Gambar 1. 1 yang merupakan data total statistik pertumbuhan sertifikat

halal di Indonesia dibandingkan dengan jumlah UMKM di Indonesia per tahun 2025, sebagai berikut.

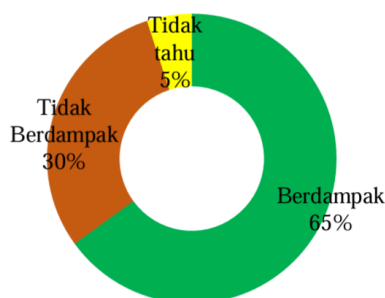


Gambar 1. 1 Data perbandingan total UMKM di Indonesia dan UMKM bersertifikasi halal 2025

Sumber: Databoks/KSP dan BPJPH

Dari adanya dari total UMKM di Indonesia dan UMKM yang bersertifikasi halal, meskipun data mengalami peningkatan tetapi jumlah UMKM bersertifikasi masih jauh dibandingkan dengan jumlah UMKM yang ada di Indonesia, sehingga hal tersebut memiliki kesenjangan atau gap permasalahan. Meskipun demikian, kebijakan ini memiliki dampak yang cukup baik terhadap pertumbuhan dari tahun 2020 hingga 2025, dan memiliki peran penting terhadap minat beli konsumen, khususnya untuk masyarakat mayoritas Muslim, yang berkontribusi banyak pada peningkatan penjualan UMKM, serta sebagai faktor yang menguatkan dampak adanya sertifikasi halal terhadap keputusan pembeli yaitu kualitas dan harga (Muntholip et al., 2025a). Dari adanya hal tersebut menunjukkan bahwa sertifikasi halal tidak hanya sebagai standar kehalalan tetapi juga sebagai strategi pemasaran yang efektif, dibuktikan dengan perlindungan konsumen yang memberikan kepastian halal bagi 87% populasi umat Muslim, sehingga kepercayaan dan minat beli meningkat hingga 30 – 50% (Muntholip et al., 2025b). dari pertumbuhan penerapan kebijakan menurut databoks dan BPJPH tersebut, Sertifikasi halal juga meningkatkan penjualan hingga 20 – 40%, akses internasional, hingga kepuasan pelanggan, terutama untuk makanan dan minuman (Larasati et al., 2023) dan mendorong ekonomi nasional industri halal senilai Rp1.287,65 triliun potensial (Fathoni, 2021).

Namun, di sisi lain, masih terdapat kesenjangan yang cukup besar antara pertumbuhan UMKM bersertifikasi dengan total UMKM yang ada di Indonesia, hal ini dapat terjadi jika pernyataan yang berbeda yang mungkin memengaruhi persepsi atau opini publik dari adanya sertifikasi ini seperti penelitian yang dilakukan oleh Roesmara Donna et al., 2025. Dapat dilihat pada Gambar 1. 2 di bawah ini.



Gambar 1. 2 Dampak sertifikasi penjualan terhadap omzet penjualan

Sumber: Roesmara Donna et al., 2025

Dari sudut pandang produsen atau UMKM dalam penelitian ini sebanyak 35% menyatakan sertifikasi halal tidak mmemengaruhi omzet penjualan mereka, sementara 5% lainnya tidak yakin, ketidakyakinan ini disebabkan oleh ketidaktahuan yang mana mereka melakukan perizinan dan pembuatan sertifikat terlebih dahulu, jadi tidak mengetahui omzet sebelum tindakan tersebut. Dan 65% produsen menyatakan bahwa sertifikasi halal menunjukkan dampak terhadap penjualan mereka. (Roesmara Donna et al., 2025). Serta dari sisi konsumen, sebagian 20% konsumen tetap memperhatikan sertifikasi halal tetapi tidak menjadikan faktor dalam keputusan pembelian mereka, hal ini menunjukkan bahwa perhatian konsumen tidak selalu tertuju pada pembelian berdasarkan sertifikasi halal. Sehingga dari hasil ini, dapat diketahui persepsi atau opini publik terkait kebijakan sertifikasi halal ini.

Dari adanya kesenjangan antara dari total jumlah UMKM di Indonesia dan UMKM yang bersertifikasi halal, menunjukkan bahwa penerapan kebijakan ini masih mengalami berbagai tantangan. Salah satunya seperti data Gambar 1. 2 yang dijelaskan di paragraf sebelumnya yaitu berkaitan persepsi masyarakat terhadap kebijakan sertifikasi halal ini yang dilakukan oleh pemerintah, dengan perbedaan pandangan dari pelaku usaha maupun konsumen sehingga dapat memengaruhi tingkat penerimaan kebijakan tersebut di kalangan masyarakat.

Seiring dengan perubahan dan berkembangnya teknologi digital, opini publik atau persepsi masyarakat terhadap suatu kebijakan telah banyak diutarakan melalui media sosial. Karena, media sosial menjadi tempat bagi masyarakat agar bisa mengekspresikan pendapat, memberikan tanggapan, maupun melakukan diskusi terkait isu publik, termasuk salah satunya isu kebijakan sertifikasi halal.

Belakangan ini, terdapat berbagai media populer salah satunya platform yang sedang memiliki tingkat interaksi tinggi, yaitu TikTok. TikTok merupakan platform media sosial yang menjadi sorotan karena bisa membuat video kreatif dengan efek unik dan menarik perhatian banyak orang, serta popularitasnya karena adanya fitur inovatif yang digunakan untuk mendorong interaksi sosial dan konten edukasi maupun informasi yang bermanfaat (Arif Efendi et al., 2025). Tingginya popularitas dan interaksi jumlah pengguna serta berbagai kegiatan yang bisa terjadi di kolom komentar tiap video kontennya, menjadikan TikTok memiliki sumber data potensial yang dapat digunakan peran sebagai media sosial yang berpengaruh untuk melihat opini publik secara luas. TikTok memberikan peran yang sangat besar akhir-akhir ini terhadap opini publik atas konten-konten yang tersedia dari berbagai user atau pengguna tiktok. Terutama pada konten mengenai topik kebijakan sertifikasi halal yang banyak dibahas dalam bentuk video konten yang memicu munculnya berbagai opini dan tanggapan masyarakat, karakteristik tersebut menjadikan TikTok sebagai sumber data yang relevan untuk menggambarkan persepsi publik terhadap kebijakan sertifikasi halal, seperti pemilihan konten dari user @fathianhafiz dan user @modeonid.media, dilakukan karena kedua konten tersebut secara spesifik membahas isu kebijakan sertifikasi halal di Indonesia serta jumlah interaksi dan komentar yang cukup tinggi, dengan pemilihan akun @fathianhaviz merupakan konten kreator yang juga seorang buruh pabrik dan aktif menyampaikan kritik, pembahasan, serta tanggapan berbagai kebijakan pemerintah melalui konten-kontennya. Sementara itu, akun @medeoind.media merupakan akun media informasi yang menyajikan pemberitaan dan isu-isu terkini yang berkembang di masyarakat. Kedua akun tersebut dipilih karena mampu memunculkan beragam opini publik terkait kebijakan sertifikasi halal, sehingga relevan sebagai sumber data dalam penelitian ini. Serta, konten TikTok yang menyebarkan informasi akan menjadi viral atau *trending topic* jika topik yang dibicarakan sangat relevan sehingga akan memengaruhi Algoritma TikTok, yang akan meningkatkan engagement seperti jumlah views, likes, dan shares sehingga akan memengaruhi opini publik dari berbagai macam pengguna TikTok yang tidak terstruktur

dan dalam jumlah yang besar pada fitur kolom komentar TikTok. Fitur ini memberikan kontribusi terhadap sifat interaktif platform ini dan mendorong keterlibatan pengguna dan memungkinkan audiens untuk meninggalkan komentar, sehingga platform ini memberikan fasilitas diskusi bagi audiens (Syahrani, 2025).

Oleh karena itu, penelitian ini dilakukan karena analisis terhadap komentar user TikTok dapat digunakan dalam menggali informasi dan mengetahui kecenderungan sentimen masyarakat terhadap kebijakan sertifikasi halal di Indonesia. Dari adanya komentar atau opini publik yang sangat banyak jumlahnya dan tidak terstruktur, memerlukan *Text Mining* dengan pendekatan *Lexicon-based Analysis* sebagai solusi pendekatan yang efektif untuk mengekstraksi informasi berupa data teks komentar TikTok dan pendekatan *Clustering* untuk pengelompokan opini yang diinterpretasikan sehingga mendapat kecenderungan sentimen publik. Nantinya setiap *cluster* akan menghasilkan sentimen positif, negatif, maupun netral serta persentase sentimen. Harapannya hasil dari penelitian ini dapat digunakan oleh pemerintah terkait agar dapat memberikan strategi yang tepat bagi publik untuk berdasarkan kelompok yang dihasilkan dari *clustering* sentimen analisis tersebut.

1.2 Rumusan Masalah

Observasi awal pada penelitian ini, ditemukan bahwa kebijakan sertifikasi halal di Indonesia menimbulkan berbagai tanggapan dari masyarakat yang disampaikan melalui media sosial TikTok. Beragamnya opini inilah yang menunjukkan perlunya analisis untuk identifikasi kecenderungan sentimen serta pola pengelompokan komentar user. Namun, dari adanya hal tersebut, perlu diketahui bagaimana distribusi sentimen dan karakteristik kelompok komentar yang terbentuk terkait kebijakan tersebut. Oleh karena itu peneliti ini dirumuskan dalam beberapa pertanyaan sebagai berikut:

1. Bagaimana hasil pengelompokan komentar konsumen Indonesia terhadap kewajiban sertifikasi halal di platform media sosial TikTok menggunakan pendekatan *lexicon-based analysis* dan *clustering*?
2. Bagaimana label dan proporsi sentimen yang terbentuk dari setiap *cluster* komentar konsumen Indonesia?
3. Bagaimana desain usulan program percepatan sertifikasi halal terhadap pelaku usaha berdasarkan kata kunci yang muncul pada *clusterisasi* opini konsumen Indonesia terhadap kewajiban sertifikasi halal?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disusun, penelitian ini bertujuan untuk menganalisis tanggapan masyarakat terhadap kebijakan sertifikasi halal di Indonesia melalui komentar user TikTok:

1. Mengelompokkan komentar konsumen Indonesia terhadap kewajiban sertifikasi halal di platform media sosial TikTok menggunakan pendekatan *lexicon-based analysis* dan *clustering*.
2. Mengidentifikasi sentimen konsumen dan mengetahui label sentimen dari *cluster* yang dihasilkan.
3. Membuat desain usulan program percepatan sertifikasi halal terhadap pelaku usaha berdasarkan kata kunci yang muncul pada *clusterisasi* opini konsumen Indonesia terhadap kewajiban sertifikasi halal.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis bagi berbagai pihak yang berkepentingan:

1. Bagi Pemerintah
Bagi pemerintah Indonesia, hasil penelitian ini dapat digunakan sebagai bahan evaluasi kebijakan dan menentukan strategi yang tepat untuk pembentukan sosialisasi terhadap masing-masing kelompok konsumen atas kebijakan sertifikasi halal.
2. Bagi Pelaku Usaha
Bagi pelaku usaha atau produsen, hasil penelitian ini dapat dipakai sebagai bahan pertimbangan dalam memahami persepsi konsumen terhadap kewajiban sertifikasi halal dan memahami aspek yang menjadi poin beban atau keuntungan, dan bahan dalam pengambilan keputusan bisnis.
3. Bagi Akademisi
Bagi akademisi, penelitian ini dapat digunakan sebagai referensi pendekatan *lexicon-based analysis* pada *Text Mining* dan *clustering* hingga implementasi *clustering* sebagai analisis sentimen agar data yang digunakan bisa dikembangkan lebih dalam.

1.5 Batasan Penelitian

Agar penelitian tetap terfokus pada permasalahan yang dikaji, maka penelitian ini dibatasi pada beberapa hal, sebagai berikut:

1. Penelitian ini hanya membahas persepsi masyarakat terhadap kebijakan sertifikasi halal di Indonesia berdasarkan komentar pengguna TikTok.
2. Analisis data dilakukan menggunakan pendekatan Text Mining dengan metode TF-IDF, *K-Means Clustering*, dan *Lexicon-Based Analysis*.
3. Penelitian ini tidak membahas dampak ekonomi, hukum maupun implementasi kebijakan sertifikasi halal secara langsung, tetapi hanya berfokus pada analisis opini yang disampaikan melalui komentar TikTok yang membahas sertifikasi halal.

1.6 Sistematika Penulisan

Penulisan penelitian ini menggunakan sistematika penulisan secara berurutan agar pembaca memahami penyusunan laporan dengan mudah, dimulai dari identifikasi masalah hingga pengusulan strategi, yang meliputi:

BAB I PENDAHULUAN

Bab ini menguraikan gambaran umum penelitian yang meliputi latar belakang, perumusan masalah, tujuan yang dicapai, manfaat penelitian, serta sistematika penulisan yang digunakan dalam penyusunan laporan penelitian.

BAB II KAJIAN PUSTAKA

Bab ini menyajikan landasan teori yang menjadi dasar pelaksanaan penelitian. Kajian pustaka disusun berdasarkan berbagai sumber referensi, seperti buku, artikel ilmiah, jurnal, dan penelitian terdahulu yang memiliki keterkaitan dengan topik yang diteliti. Selain itu, bab ini juga memaparkan konsep-konsep yang relevan untuk mendukung proses analisis dan pengolahan data dalam penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode dan tahapan penelitian yang dilakukan secara sistematis. Uraian mencakup proses pengumpulan data, pengolahan data, hingga teknik analisis yang diterapkan dalam penelitian. Serta, alur penelitian disajikan dalam bentuk diagram untuk memberikan gambaran yang jelas mengenai setiap tahapan yang dilalui, seperti tahapan penggunaan algoritma TF-IDF, K-Means, metode Elbow dalam analisis komentar oleh user TikTok.

BAB IV HASIL DAN PENGOLAHAN DATA

Bab ini memaparkan hasil pengolahan data yang diperoleh dari komentar suatu konten TikTok. Selain itu, disajikan hasil analisis clusterisasi beserta interpretasi terhadap data yang telah diolah. Penyajian hasil dilakukan melalui tabel, grafik, dan visualisasi seperti *wordcloud* untuk mendukung pemahaman terhadap temuan penelitian.

BAB V PEMBAHASAN

Bab ini membahas secara rinci hasil analisis yang telah diperoleh pada tahap sebelumnya. Pembahasan difokuskan pada interpretasi temuan penelitian, identifikasi pola yang muncul dari hasil klasterisasi, serta penyusunan strategi untuk percepatan sertifikasi halal industri di Indonesia.

BAB VI KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang disusun berdasarkan hasil penelitian dan ditujukan untuk menjawab rumusan masalah yang telah diterapkan. Selain itu, bab ini juga memuat rangkuman temuan utama penelitian serta saran yang dapat digunakan sebagai bahan pertimbangan dalam penelitian selanjutnya.

DAFTAR PUSTAKA**LAMPIRAN**

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Literatur

Kajian literatur yang digunakan pada penelitian ini mencakup analisis sentimen terhadap topik kewajiban sertifikasi halal berdasarkan opini atau komentar melalui media sosial seperti TikTok, Instagram, Twitter, dan objek penelitian lainnya seperti platform E-Commerce. Misalnya pada penelitian terdahulu yang dilakukan oleh (Hakim et al., 2022), penelitian ini berfokus pada analisis sentimen publik terhadap sertifikasi halal sebagai objek utama nya melalui media sosial Twitter sebagai sumber data opini publik untuk diteliti dalam konteks ekonomi. Penelitian lain, terdapat penelitian yang dilakukan oleh (Timur et al., 2023), dengan topik yang berfokus pada sentimen dan opini netizen Indonesia mengenai sertifikasi halal sebagai objek utama penelitian berbasis media sosial Twitter yang dianalisis pada *Text Mining* dengan topic pemodelan pada Orange Data Mining menggunakan metode *WordCloud*.

Pada saat memperoleh hasil analisis sentimen beberapa peneliti terdahulu menggunakan metode yang berbeda-beda misalnya, Penelitian yang dilakukan oleh (Jayus et al., 2024), analisis dilakukan pada kepuasan konsumen terhadap produk halal dengan menggunakan metode *Latent Dirichlet Allocation* (LDA) dan VADER. Peneliti lain yaitu (Nury et al., 2024) melakukan penelitian terhadap es Krim Mixue yang tadinya belum mendapat sertifikasi halal, hal ini menyebabkan keraguan masyarakat terhadap Es Krim tersebut, peneliti melakukan penelitian melalui media sosial Twitter dari komentar pada akun @CNNIndonesia, kemudian mengolahnya menggunakan tools NoLimit dan memperoleh hasil analisis sentimen negatif sebanyak 53,3%, netral 46,7%, dan positif 0% pada 28 – 30 Desember 2022, dan pada 16 – 20 Februari 2023 diperoleh netral sebanyak 63,7%, negatif 21,2%, dan positif 15,0%. Penelitian lain yaitu dilakukan oleh (Kasturi et al., 2025), dengan maksud melakukan penelitian ini untuk melihat apakah ada pengaruh minat beli mahasiswa Fakultas Ekonomi Islam Universitas Islam Jakarta terhadap sertifikasi halal dari media sosial seperti TikTok menggunakan metode kuesioner dan SmartPLS 4.0.

Penelitian selanjutnya yaitu (Abidin & Niskaromah, 2025) mengenai pengaruh media sosial terhadap kesadaran masyarakat Indonesia mengenai sertifikasi halal pada warga Kampung Laut, Kuala Jambi, Indonesia, penelitian ini mengidentifikasi paparan konten terkait

sertifikasi halal pada platform Media sosial dengan metode pengumpulan data dengan penyebaran kuesioner, responden sebanyak 250 responden, menggunakan regresi linier sederhana, dengan hasil 65% responden pernah melihat atau membaca konten terkait sertifikasi halal di media sosial, 53% responden memperoleh pengetahuan baru, 60% responden memiliki sikap positif dan kepercayaan terhadap konten sertifikasi halal di media sosial. Berikutnya, terdapat penelitian oleh (Bakti & Ratnasih, 2025), yang melakukan penelitian terhadap persepsi customer terhadap digital halal sertifikasi dan implikasi nya untuk produk MSME Purchasing, yang diperoleh hasil 73% responden memiliki persepsi positif melalui metode regresi dan *Confirmatory Factor Analysis*. Penelitian selanjutnya mengenai analisis sentimen terhadap 88 artikel Scopus tentang halal food, yang mana penelitian yang dilakukan oleh (Izza et al., 2022) ini melakukan analisis sentimen secara Meta-Analysis dan memperoleh hasil utama positif sebanyak 43,2%, netral 36,4%, dan Negatif 19,3% dengan mayoritas menunjukkan dukungan terhadap halal food (sertifikasi halal). Selanjutnya, terdapat penelitian oleh (Rahel et al., 2025), yang melakukan penelitian terhadap keputusan konsumen Muslim di TikTok Shop yang dipengaruhi oleh promosi dan sertifikasi halal, dengan pengumpulan data berupa kuesioner skala Likert, dan dianalisis menggunakan metode Regresi Linier berganda, hasil yang diperoleh bahwa sertifikasi halal ini sebagai aspek yang paling dominan berpengaruh terhadap konsumen Muslim. Penelitian berikutnya dilakukan oleh (Masrukhan & Ifrizah, 2024), dengan meneliti analisis sentimen berfokus pada media sosial seperti Twitter, Instagram, Facebook, untuk memahami persepsi konsumen terhadap produk halal bersertifikasi. Hasil dari penelitian ini 60% menunjukkan sentimen positif, 25% menunjukkan sentimen netral, dan 15% menunjukkan sentimen negatif. Penelitian selanjutnya oleh (Sari & Rahayu, 2024) menerapkan pendekatan kualitatif dan analisis sentimen menggunakan metode polaritas teks dengan SentiStrength, yang meneliti tentang literatur akademik (*halal traceability*) pada periode 2004 hingga 2023, yang menunjukkan bahwa hasil analisis sentimen didominasi oleh sentimen netral yang masih memiliki kekhawatiran akan pemahaman sertifikasi halal serta pemanfaatan teknologi *Blockchain*. Berikutnya terdapat penelitian oleh (Hardimanto, 2025), yang meneliti menggunakan metode regresi linier sederhana untuk analisis keputusan pembelian konsumen daerah Surakarta melalui Kuissoner terhadap pelabelan halal, yang menghasilkan dampak positif dengan adanya sertifikasi halal terutama komunitas Muslim. Penelitian berikutnya dilakukan oleh (Nanta Setia Dien Alfansyuri & Arsyi, 2023), penelitian ini dilakukan dengan menganalisis minat beli konsumen Muslim dari adanya pengaruh label halal pada packaging produk, dengan

menggunakan metode regresi linier sederhana didapat hasil bahwa konsumen rata-rata setuju dan paham tentang indikator pentingnya halal dan label halal pada produk. Penelitian lainnya yaitu oleh (Nurwijayanto et al., 2024), memiliki persamaan dengan penelitian ini terkait topik yang diangkat mengenai sertifikasi halal dengan metode regresi linier berganda dengan objek konsumen pada Angkringan Tenda Mak Gendut, yang mana sertifikasi halal menjadi dominan pada faktor yang memengaruhi keputusan pembelian konsumen di UMKM tersebut. Penelitian terakhir dilakukan oleh (Afandi et al., 2025), dengan analisis sentimen terhadap kebijakan sertifikasi halal terhadap opini publik di Twitter menggunakan Metoded Naïve bayes dan KNN, dengan hasil perolehan sentimen sebanyak 57,3% netral, 21,1% positif, dan 15,7% negatif.

2.1.1 Posisi Penelitian

Tabel 2. 1 di bawah ini merupakan informasi mengenai posisi penelitian terdahulu dengan penelitian sekarang yang akan dilakukan:

Tabel 2. 1 Posisi Penelitian

No	Penulis	Objek Penelitian				Algoritma Penelitian					
		Tik Tok	Twitter/X	Instagram	Lainnya	TF-IDF	Naïve Bayes/ (SVM)	Vader/ Lexicon Based	K-Means	Chi-square	Lainnya Clustering
1	(Hakim et al., 2022)							✓			
2	(Erlangga & Patimban gi, 2025)		✓		✓						

No	Penulis	Objek Penelitian				Algoritma Penelitian						
		Tik Tok	Twitter/ X	Instagram	Lainnya	TF-IDF	Naïve Bayes/ (SVM)	Vader/ Lexicon Based	K-Means	Chi-square	Lainnya	Clustering
3	(Jayus et al., 2024)		✓					✓			✓	✓
4	(Astiwara, 2025)				✓					✓		
5			✓									
6	(Abidin & Niskaromah, 2025)	✓		✓								
7	(Bakti & Ratnasih, 2025)				✓			✓				
8	(Izza et al., 2022)				✓				✓		✓	
9	(Rahel et al., 2025)	✓				✓			✓		✓	
10	(Masrukh an & Ifrizah, 2024)		✓	✓	✓		✓					

No	Penulis	Objek Penelitian				Algoritma Penelitian						
		Tik Tok	Twitter/X	Instagram	Lainnya	TF-IDF	Naïve Bayes/ (SVM)	Vader/ Lexicon Based	K-Means	Chi-square	Lainnya	Clustering
11	(Sari & Rahayu, 2024)				✓						✓	
12	(Hardianto, 2025)				✓						✓	
13	(Nanta Setia Dien Alfansyuri & Arsyi, 2023)				✓						✓	
14	(Nurwijayanto et al., 2024)				✓						✓	
15	(Afandi et al., 2025)		✓				✓					

No	Penulis	Objek Penelitian				Algoritma Penelitian						
		Tik Tok	Twitter/X	Instagram	Lainnya	TF-IDF	Naïve Bayes/ (SVM)	Vader/ Lexicon Based	K-Means	Chi-square	Lainnya	Clustering
16	Penelitian Selanjutnya	✓				✓			✓			✓

Dari adanya penelitian terdahulu di atas, terdapat persamaan pada masing-masing penelitian dengan penelitian yang dilakukan ini, topik kewajiban sertifikasi halal telah menunjukkan tanggapan publik atau konsumen dengan berbagai reaksi, serta peneliti terdahulu melakukan analisis terhadap topik tersebut dengan karakteristik dan metode yang berbeda-beda, sebagai penelitian baru yang dilakukan ini, terdapat hal yang berbeda dari penelitian ini atau gap penelitian yaitu dapat dilihat dari segi metode penelitian, sumber data, dan objek penelitian. Misalnya, Pada penelitian terdahulu yang dilakukan oleh (Hakim et al., 2022), penelitian ini menunjukkan pada analisis sentimen publik terhadap sertifikasi halal sebagai objek utama nya melalui media sosial Twitter sebagai sumber data opini publik untuk diteliti dalam konteks ekonomi, dengan analisis *Text Mining* yaitu menggunakan metode VADER dan *Lexicon Based Analysis*, sedangkan pada penelitian ini tidak menggunakan pendekatan VADER tetapi masih menggunakan *Lexicon Based Analysis* tersebut.

Di sisi lain, penelitian ini menggunakan pendekatan gabungan yaitu *Lexicon-based Analysis* dengan pembobotan algoritma TF-IDF (*Term Frequency – Inverse Document Frequency*) dan *Clustering* dengan algoritma K-Means, dengan objek penelitian dari opini publik melalui media sosial TikTok, seperti pada penelitian terdahulu oleh (Kasturi et al., 2025), (Abidin & Niskaromah, 2025), dan (Rahel et al., 2025). Selanjutnya terdapat penelitian yang dilakukan oleh (Jayus et al., 2024), yang melakukan analisis pada kepuasan konsumen terhadap produk halal dengan menggunakan metode *clustering* juga, tetapi metode analisis sentimen digunakan berbeda yaitu menggunakan pemodelan topik *Latent Dirichlet Allocation* (LDA) dan VADER. Sedangkan pada penelitian ini menggunakan algoritma TF-IDF (*Term Frequency – Inverse Document Frequency*) dengan *tools* Google Colab.

2.2 Landasan Teori

2.2.1 Kebijakan Sertifikasi Halal

Sertifikasi Halal adalah suatu proses untuk memperoleh sertifikat halal melalui beberapa tahap untuk membuktikan bahwa bahan, proses produksi dan SJH (Sistem Jaminan Halal) memenuhi standar LPPOM MUI (LPPOM MUI, 2008). Menurut Pasal 1 Ayat 10 Undang-Undang Nomor 33 Tahun 2014 tentang Jaminan Produk Halal, memiliki definisi terhadap Sertifikasi halal yang merupakan pengakuan tertulis yang diterbitkan oleh BPJPH (Badan Penyelenggara Produk Halal) atas Produk Halal berdasarkan fatwa MUI (Majelis Ulama Indonesia). Undang-Undang ini disahkan pada 17 Oktober 2014, dan amandemen melalui UU No. 11/2020 tentang Cipta

Kerja. Undang-Undang dan definisi ini dibuat untuk melindungi umat Muslim terhadap jaminan produk halal di Indonesia.

Kewajiban sertifikasi halal diberlakukan mulai 17 Oktober 2019, kewenangan berada di bawah BPJPH (Badan Penyelenggara Jaminan Produk Halal) Kementerian Agama. BPJPH dapat membentuk perwakilan di daerah, dengan ketentuan fungsi, tugas, dan susunan organisasi BPJPH diatur dalam Peraturan Presiden. Lembaga berwenang merupuskan dan menetapkan kebijakan JPH, termasuk menetapkan norma, standar, prosedur, dan kriteria JPH, serta menerbitkan dan mencabut sertifikat halal dan label halal pada produk (CNN Indonesia, 2019). Adanya kebijakan sertifikasi halal ini memiliki fungsi sebagai standarisasi halal yaitu memberikan kepastian, perlindungan dan ketenangan konsumen, terutama umat Islam, dari mengkonsumsi suatu produk yang haram, hal ini terdapat dalam Undang-Undang Nomor 8 Tahun 1999 tentang Perlindungan Konsumen. Tujuan dilakukannya sertifikasi halal agar perusahaan memberikan jaminan halal dengan selalu menjaga konsistensi kehalalan produk dengan melakukan rangkaian pemeriksaan yang dilakukan oleh auditor yang ahli di bidangnya agar dapat dinyatakan status kehalalannya sehingga tertulis fatwa halal pada suatu produk sertifikat halal (Hasan, 2014).

2.2.2 TikTok

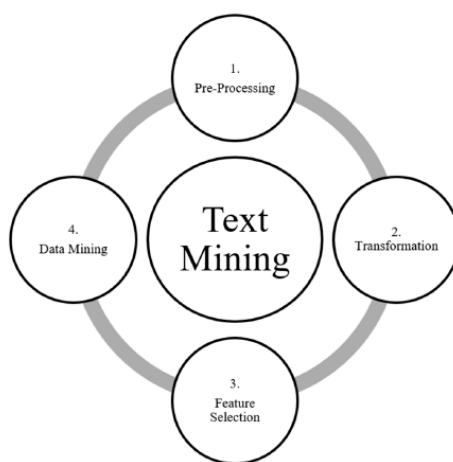
TikTok adalah salah satu platform media sosial yang memungkinkan penggunanya untuk membuat dan membagikan video pendek dengan durasi maksimal 3 menit yang dirancang untuk mempromosikan konten viral. TikTok telah mengubah cara pandang orang mengkonsumsi dan memproduksi konten secara digital. TikTok pertama kali ada pada tahun 2016 oleh perusahaan China yaitu ByteDance (Panayitsa & Al Falah, 2024). TikTok merupakan sebuah aplikasi yang memberikan efek spesial yang unik, menarik, dan bisa digunakan oleh para pengguna aplikasi ini dengan mudah untuk membuat video pendek yang keren dan bisa menarik perhatian banyak orang yang melihatnya (Malimbe et al., 2021).

TikTok tidak hanya memberikan konten hiburan tetapi juga menjadi platform yang bisa digunakan sebagai wadah untuk menyalurkan berita dan informasi. Penggunaan TikTok dalam konteks berita mencerminkan pergeseran pola konsumsi informasi masyarakat yang semakin mengarah ke platform digital, dengan konten yang mampu menarik perhatian secara visual dan interaktif (Isnaina Akbari et al., 2025).

TikTok menjadi aplikasi paling banyak diminati sebanyak 45,8 juta kali unduhan, dan dari banyaknya pengguna aktif TikTok sebesar 625 juta menjadikan platform ini sebagai sarana pemberian informasi yang cepat dan menarik (Rahmana et al., 2022). Sistem FYP (For You Page) menampilkan berita saat kita sedang tidak ingin baca berita lebih lanjut, namun adanya sistem pencarian seperti google, fitur komentar, dan duet yang memungkinkan diskusi dan penyebaran informasi berita yang lebih interaktif. Peran fitur komentar ini salah satunya memiliki aksesibilitas dan interaktivitas seperti fitur komen dan kolom pencarian yang dirancang mirip google, yang memungkinkan diskusi dan penyebaran informasi lebih interaktif. Hal ini dianggap menarik dan memberikan dampak karena komentar-komentar di TikTok dapat memberi tahu apakah isi video tersebut informatif, menarik, atau tidak berguna sama sekali (Parapat & Azhar, 2024).

2.2.3 Text Mining

Text Mining merupakan proses untuk mengekstraksi informasi dari sumber data yang tidak terstruktur dengan menggunakan teknik penambangan data guna menganalisis dan mengelola data (Putri & Hendrowati, 2018). *Text Mining* memiliki kombinasi teknik seperti *data mining*, *machine learning*, *natural language processing*, *information retrieval*, dan *knowledge management* (Widaningsih & Suheri, 2018). Menurut (Chowdhury & Alzarrad, 2023) terdapat proses yang dilakukan selama tahapan *Text Mining* yaitu.



Gambar 2. 1 Proses *Text Mining*

Sumber: (Chowdhury & Alzarrad, 2023)

Dari Gambar 2. 1 Pada tahap (*pra-processing*) dilakukan dengan mengubah teks menjadi bentuk yang dapat diprediksi dan dianalisis untuk tugas penambangan teks (*data mining*), data

teks pada umumnya mengandung noise sehingga dilakukan pembersihan terlebih dahulu melalui tahapan ini. Lalu tahap (*transformation*), pada tahap ini melibatkan representasi dokumen berdasarkan teks yang dikandungnya dan jumlah kemunculannya, dengan dua metode yaitu model ruang vektor dan model *bag of words* (BoW). Selanjutnya (*feature selection*), pemilihan fitur atau atribut atau variabel, dengan memilih fitur yang paling relevan dari variabel yang tersedia yang mengurangi kesalahan prediksi. Terakhir, terdapat (*data mining*), pada proses ini akan dilakukan analisis data dan menghasilkan wawasan yang berguna, teknik penambangan data ini mencakup klasifikasi, pengelompokan (*clustering*), regresi, *association of mining*, dan banyak lagi.

Dalam *Text Mining*, sebelum melakukan analisis data dilakukan terlebih dahulu langkah yang disebut *pre-processing* seperti yang telah disebutkan di atas, yang bertujuan untuk mengubah teks dari awalnya tidak terstruktur menjadi lebih terstruktur sehingga teks dapat diolah lebih lanjut (Putri & Hendrowati, 2018). Menurut (Afrad et al., 2024) tahapan *pre-processing* diantaranya sebagai berikut:

1. *Case Folding*

Pada *Text Mining*, Proses ini dilakukan saat teks yang terdiri dari berbagai huruf besar dan kecil distandarisasi. Contoh kata sebelum dan sesudah dilakukan *case folding* yaitu ‘Tempat’ menjadi ‘tempat,’ dan ‘Tahun’ menjadi ‘tahun.’

2. *Tokenize*

Tahap selanjutnya, terdapat tokenize atau tokenisasi yaitu proses saat teks kalimat dipecah menjadi kata atau token. Langkah ini sangat penting untuk analisis teks karena mengubah teks yang tidak terstruktur menjadi format terstruktur yang lebih cepat diproses oleh algoritma. Contoh proses tokenisasi yaitu [‘tempat wisata yang sejuk dan menyenangkan’] menjadi [‘tempat,’ ‘wisata,’ ‘yang,’ ‘sejuk,’ ‘dan,’ ‘menyenangkan.’]

3. *Stopword Removal*

Pada tahap ini, kata-kata yang dianggap tidak bermakna dan sering muncul dalam jumlah yang besar dalam data akan dihilangkan. Contoh kata yang dihilangkan yaitu ‘dan,’ ‘atau,’ ‘yang.’

4. *Stemming*

Stemming digunakan untuk mengubah kata menjadi bentuk akarnya. Contoh dalam penelitian (Afrad et al., 2024) menggunakan perpustakaan sastra yang memiliki

kamus akar kata Indonesia, contohnya proses *stemming* dari kata ‘menyenangkan’ menjadi kata dasar ‘senang.’

2.2.4 Analisis Sentimen

Analisis sentimen atau biasa disebut juga *opinion mining* merupakan kajian yang menganalisis opini, sentimen, penilaian, sikap, dan emosi manusia terhadap suatu entitas dan atribut yang diekspresikan dalam teks tertulis. Entitas di sini maksudnya seperti produk, layanan, organisasi, individu, peristiwa, isu, atau topik (Liu, 2020). Menurut kamus Merriam – Webster, definisi *sentiment* adalah sikap, pemikiran, atau penilaian yang dipicu oleh perasaan, sedangkan definisi dari *opinion* adalah pandangan, penilaian, atau evaluasi yang terbentuk dalam pikiran mengenai suatu hal tertentu. Perbedaan tersebut menunjukkan bahwa sentimen lebih mengedepankan perasaan sedangkan opini bersifat konkret terhadap sesuatu. (Liu, 2020).

Sentimen memiliki jenis diantaranya dapat berupa sentimen positif, netral, negatif, yang berikutnya dapat dibagi lagi ke dalam emosi seperti keterkejutan (*surprise*), kepercayaan (*trust*), antisipasi (*anticipation*), kemarahan (*anger*), ketakutan (*fear*), kesedihan (*sadness*), jijik (*disgust*), kebahagiaan (*joy*), dan lainnya (Mao et al., 2024). Selain jenis sentimen, terdapat implementasi dari analisis sentimen yaitu pada *Government intelligence*, dari pemantauan berbagai sumber, dan meningkatkan komunikasi yang bersifat antagonistik atau bermusuhan dapat dilacak. Biasanya untuk mendukung keputusan dalam perumusan kebijakan yang efektif, analisis sentimen dapat digunakan untuk membantu analisis otomatis opini masyarakat terhadap kebijakan yang sedang direncanakan atau usulan regulasi pemerintah. Seperti menerapkan pelacakan opini masyarakat terhadap program baru dan lainnya (Kumar & Sebastian, 2012).

2.2.5 Pembobotan TF-IDF

Metode pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah skema pembobotan yang digunakan untuk menghilangkan istilah dengan bobot lebih rendah dari dokumen dan membantu meningkatkan efektivitas pengambilan. frekuensi dokumen terbalik (IDF) adalah statistik numerik yang memberi tahu kita seberapa penting suatu kata bagi suatu dokumen dalam koleksi atau korpus. Dan sebagian besar digunakan untuk pengambilan informasi dan *Text Mining* (Kumari et al., 2016). Jadi tujuan metode ini adalah untuk evaluasi seberapa penting suatu *term* (kata) dalam sebuah dokumen (TF). Adapun rumus penggunaan TF-IDF sebagai berikut.

1. *Term Frequency* (TF)

TF berfungsi untuk Mengukur seberapa sering kata (term) muncul dalam sebuah dokumen. Rumus menurut (Septiani & Isabela, 2022) adalah sebagai berikut:

$$Tf = 0,5 + 0,5 \frac{tf}{\max(tf)} \quad (2.1)$$

Jadi, rumus di atas dilakukan dengan menghitung jumlah kata tersebut dibagi dengan jumlah total kata dalam dokumen

2. Inverse Document Frequency (IDF)

IDF digunakan untuk mengukur seberapa penting suatu kata dalam konteks koleksi dokumen yang lebih besar. IDF dihitung dengan membagi jumlah total dokumen dalam koleksi dengan jumlah dokumen yang mengandung kata tersebut. Hasilnya diambil log untuk memperhalus skala. Berikut rumus IDF menurut (Septiani & Isabela, 2022).

$$idf_t = \log \frac{tD}{df_t} \quad (2.2)$$

$$W_{d.t} = tf_{d.t} \times idf_{d.t} \quad (2.3)$$

Dalam metode TF- IDF, nilai hasil perkalian antara TF-IDF akan menghasilkan bobot kata (*term weight*) untuk setiap kata dalam sebuah dokumen. Hasil dari ini mencerminkan tingkat seberapa pentingnya kata dalam dokumen tersebut dibanding koleksi dokumen yang lebih besar, menggunakan rumus berikut (Septiani & Isabela, 2022).

2.2.6 Clustering

Clustering atau klasterisasi merupakan metode pengelompokan data. Metode yang mengelompokkan sebuah set objek data ke dalam himpunan bagian yang dikenal sebagai *cluster*, yang memungkinkan data dalam satu *cluster* memiliki tingkat kemiripan yang paling tinggi dan data di antara *cluster* memiliki tingkat kemiripan paling rendah. *Clustering* merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan *cluster* (Triayudi et al., 2024).

Menurut (Rai et al., 2010), *Clustering* adalah suatu proses menemukan kelompok objek sedemikian rupa sehingga objek-objek dalam satu kelompok akan mirip (atau berhubungan)

satu sama lain dan berbeda dari (atau tidak berhubungan dengan) objek-objek dalam kelompok lain. *Clustering* setara dengan memecah grafik menjadi komponen-komponen yang terhubung satu sama lain. *Clustering* ini memiliki tujuan dengan digabungkan dengan menilai kemiripan centroidnya berdasarkan kedekatan dalam ruang fitur, menggabungkan klaster dengan kemiripan tertinggi sesuai dengan kriteria keterkaitan yang dipilih (Wani, 2024).

2.2.7 Algoritma K-Means

Algoritma K-Means merupakan alat pengelompokan paling populer yang digunakan dalam aplikasi ilmiah dan industri. Nama tersebut berasal dari representasi setiap k klaster C dengan rata-rata (atau rata-rata tertimbang) C dari titik-titiknya, yang disebut centroid. Algoritma ini memiliki pemahaman geometris dan statistik yang baik untuk atribut numerik (Rai et al., 2010). Jumlah perbedaan antara sebuah titik dan *centroid* yang dinyatakan melalui jarak yang sesuai digunakan sebagai fungsi tujuan, setiap titik ditugaskan ke klaster dengan centroid terdekat. Dan sebelum melakukan pengelompokan jumlah klaster k harus ditentukan (Rai et al., 2010). Proses pengklasifikasian menggunakan metode K-means ini ditentukan dengan menghitung jarak objek terhadap *centroid* yaitu persamaan jarak *eucilidian*. Menurut teori yang disampaikan oleh (Rai et al., 2010) tersebut, oleh karena itu algoritma K-Means diterapkan karena setelah proses representasi menggunakan TF-IDF karena mampu mengelompokkan dokumen berdasarkan kemiripan kata yang terkandung di dalamnya. Sesuai dengan tujuan dan rancangan penelitian ini, sehingga sesuai dengan data penelitian ini yang berfokus pada identifikasi pola kelompok (*cluster*) komentar pengguna TikTok setelah dilakukan TF-IDF, maka membantu dapat diketahui pola pengelompokan pengguna TikTok terkait kebijakan sertifikasi halal berdasar tingkat kemiripan antar komentar nya. Berikut merupakan proses penentuan kelompok menggunakan *K-Means* (Mara & Intisari, 2013):

1. Menentukan jumlah *cluster* (k).
2. Mengalokasikan objek ke dalam *cluster* secara acak.

$$v = \frac{\sum_{i=1}^n x_i}{n} ; i = 1, 2, 3, \dots n \quad (2.4)$$

3. Menghitung *centroid* sampel yang ada pada masing-masing *cluster*, menggunakan *euclidean distance*.

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; i = 1, 2, 3, \dots, n \quad (2.5)$$

Keterangan: x_i : objek x ke- i
 y_i : daya y ke- i
 n : banyaknya objek

4. Mengalokasikan objek ke dalam masing-masing centroid terdekat.
5. Kemudian, kembali ke langkah 3 jika terdapat objek yang masih ada perubahan nilai *centroid*, atau di atas nilai *threshold* yang ditentukan, ataupun jika perubahan nilai *objective function* yang digunakan di atas nilai *threshold*.

Dari langkah-langkah tersebut, digunakan untuk melakukan pengecekan terhadap hasil perhitungan, apabila telah sesuai dan tidak terdapat perubahan, maka telah hasil perhitungan telah selesai, dengan mendapat *cluster* (kelas) yang diinginkan peneliti.

2.2.8 Metode Elbow

Metode Elbow ini merupakan metode yang digunakan untuk menentukan jumlah *cluster* (kelas) terbaik. Metode ini dilakukan dengan cara melihat presentase setiap *cluster* yang akan membentuk siku pada suatu titik tertentu. Metode ini disajikan dalam grafik biasanya berupa plot untuk mengetahui lebih jelas siku yang terbentuk (Maori, 2023). Cara kerja dari metode Elbow yaitu membandingkan nilai atau presentase dari sejumlah k yang telah diujikan dan membentuk siku. Nilai k pada kombinasi siku dengan *k-means* grafik hubungan *cluster* dengan penurunan error. Peningkatan nilai k menyebabkan grafik menurun perlahan sampai hasil nilai k stabil. Kemudian jumlah *cluster* k yang dihasilkan dari pengujian dengan *k-means* dievaluasi menggunakan teknik SSE sebagai berikut (Jollyta et al., 2019):

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|_2^2 \quad (2.6)$$

2.2.9 Lexicon-based Analysis

Lexicon-based Analysis merupakan pendekatan analisis berdasarkan *Lexicon* untuk klasifikasi ekspresi sentimen yang meliputi kata sentimen, frasa, idiom, dan aturan komposisi (Liu, 2020). Serta, seperangkat aturan untuk menangani berbagai konstruksi bahasa seperti *sentiment*

shifters dan klausa ‘tetapi’, dan jenis-jenis kalimat, serta target yang diturunkan untuk menentukan orientasi sentimen dalam sebuah kalimat (Ding et al., 2008 :Ding & Liu, 2010). Dari metode ini diasumsikan bahwa entitas dan target aspek telah diketahui sebelumnya, metode ini terdiri dari empat langkah yaitu:

1. Pendandaan ekspresi sentimen (kata dan frasa)

Ekspresi sentimen diketahui dalam setiap kalimat yang mengandung satu atau lebih aspek (termasuk entitas). Setiap ekspresi positif diberi skor sentimen +1, sedangkan setiap ekspresi negatif diberi skor sentimen -1. Sentimen dapat disimpulkan berdasarkan konteks.

2. Penerapan sentimen *shifters*

Sentimen *shifters* merupakan kata atau frasa yang dapat diubah orientasi sentimen. Jenis *shifter* yang umumnya digunakan yaitu *not, never, none, nobody, nowhere, neither, dan cannot*. Hal ini mengubah kalimat contoh “*The voice quality of this phone is not good [-1], but the battery life is long*” karena ada kata negasi *not* (tidak). Hal semacam ini perlu diidentifikasi berdasarkan lexicon dan dikecualikan dari proses klasifikasi sentimen.

3. Penanganan klausa *but* (tetapi)

Kata atau frasa yang memiliki pertentangan perlu penanganan khusus karena sering merubah orientasi sentimen. Kata *tetapi* ‘*but*’ dalam bahasa inggris ini mengandung pertentangan. Jadi harus memiliki aturan berikut: orientasi sentimen sebelum kata pertentangan misalnya ‘*but*’, misalnya pada kalimat “*Car-x is great, but Car-y is better.*” Setelah ini, kalimat menjadi “*The voice quality of this phone is not good [-1], but the battery life is long [+1],*” skor [+1] ditambahkan pada akhir klausa ‘*but*’. Dan dalam konteks ini disimpulkan bahwa ‘*long*’ memiliki sentimen positif terhadap aspek daya tahan baterai. Selain *but*, juga terdapat frasa lain seperti *however, with exception of, exept that, dan exept for*.

4. Agregasi skor sentimen

Langkah selanjutnya, fungsi agregasi sentimen atau opini diterapkan pada skor sentimen yang dihasilkan untuk menentukan orientasi akhir sentimen setiap aspek kalimat. Terdapat sebuah kalimat s yang mengandung sekumpulan aspek $\{a_1, \dots, a_m\}$ dan sekumpulan ekspresi sentimen $\{se_1, \dots, se_n\}$ serta skor sentimennya dari langkah 1 –

BAB III

METODE PENELITIAN

3.1 Objek dan Subjek Penelitian

Penelitian ini memiliki objek penelitian yaitu komentar yang diberikan oleh pengguna TikTok dalam suatu konten TikTok. Komentar tersebut digunakan sebagai objek untuk melakukan analisis sentimen dan mendapatkan pola kelompok komentar (kelas) dari konsumen atau komentator terhadap topik Kewajiban Sertifikasi Halal di Indonesia dengan Konsumen adalah sebagai subjek nya.

3.2 Pengumpulan Data Penelitian

Populasi yang digunakan dalam penelitian ini adalah data historis komentar suatu konten TikTok yang diambil dari dua konten yang mengusung tentang Kebijakan Sertifikasi Halal yaitu bersumber dari akun TikTok @fathianhafiz dan @medeoind.media, dengan jumlah keseluruhan sampel sebanyak ribuan komentar. Komentar ini dimulai sejak tanggal 2 Februari 2024 hingga 29 November 2024. Sehingga data yang diambil memiliki batasan hanya dari bulan Februari 2024 hingga November 2024.

Periode waktu tersebut ditentukan berdasarkan respon masyarakat yang masih memiliki ragam respon di tahun 2024 ini terhadap adanya Kebijakan Sertifikasi Halal yang sudah mulai disosialisasikan dari tahun 2019. Untuk jumlah komentar yang diambil sebagai sampel penelitian ini sebanyak 1100 komentar dari TikTok, dengan komentar berbahasa Indonesia. Penentuan serta pemilihan konten dan komentar yang diambil ini memiliki tujuan untuk menganalisis persepsi konsumen atau pengguna TikTok terhadap kebijakan sertifikasi halal ini, selain itu juga untuk mengetahui perkembangan maupun perubahan atas komentar terhadap kebijakan tersebut dari tahun ke tahun.

3.3 Metode Pengumpulan Data Penelitian

Metode pengumpulan data untuk penelitian ini menggunakan metode *comment scraping* menggunakan alat bernama *Apify*. Melalui *Apify* ini dilakukan *running* menggunakan *TikTok Comments Scraper*. Kemudian, data yang sudah diambil secara *scrapping* dari TikTok diolah menggunakan *Google Colaboratory*. *Google Colaboratory* atau biasa disebut *Google Colab* ini merupakan tools yang digunakan untuk menjalankan kode pemrograman *Phyton*, dan kode lainnya tanpa perlu melakukan instalasi terlebih dahulu.

3.4 Jenis dan Sumber Data

Jenis data yang digunakan pada penelitian ini yaitu data sekunder. Data sekunder diperoleh dari komentar melalui konetn TikTok dengan bantuan alat TikTok *Comments Scraper* dari platform *Apify*. Serta, sumber data yang digunakan yaitu berasal dari komentar yang ada konten akun TikTok yaitu dari user @fathianhafiz <https://vt.tiktok.com/ZSPFUsuwH/> dan @medeoind.media <https://vt.tiktok.com/ZSPFUTwE5/>. Hasil dari data yang diambil tersebut berupa komentar dari dua konten tersebut yang diambil sebanyak 1100 komentar dari tanggal 2 Februari 2024 hingga 29 November 2024. Selain itu, sumber data sekunder diperoleh dari berbagai jurnal, E-book, artikel, berita, dan informasi yang relevan dengan data penelitian.

3.5 Variabel Penelitian

Pada data penelitian ini, memiliki variabel yang menyajikan informasi terkait data komentar. Tetapi, di dalam penelitian ini hanya memanfaatkan tiga variabel pokok dari keseluruhan variabel, diantaranya yaitu:

a. *UniqueId*

Variabel ini digunakan sebagai petunjuk user atau pengguna yang memberikan text atau komentar opini di aplikasi TikTok.

b. *CreateTimeISO*

Variabel *CreateTimeISO* digunakan sebagai petunjuk informasi terhadap waktu kapan komentar tersebut dibuat oleh pengguna. Variabel ini digunakan sebagai landasan dalam pembatasan komentar dalam periode waktu tertentu.

c. *Text*

Variabel *Text* ini digunakan sebagai petunjuk informasi teks komentar yang dibuat oleh pengguna. Variabel ini menjadi data pokok yang digunakan dalam penelitian ini, yang nantinya akan ditelaah, diolah, dan di transformasi.

3.6 Metode Analisis Data

Metode analisis data yang digunakan dalam penelitian ini yaitu menggunakan tools Google Colab dengan bahasa *machine learning* dan *Microsoft Excel* dalam memperoleh data. Berikut merupakan metode yang digunakan sebagai analisis data komentar dalam penelitian ini:

1. *Pre-Processing Data*

Metode analisis data yang pertama dilakukan dengan pengolahan data yaitu melakukan tahap pre-processing data dengan tujuan untuk mengeliminasi atau menghapus *noise*, *outliner* data, tanda baca tidak berhubungan, yang mengubah teks dari awalnya tidak terstruktur menjadi lebih terstruktur sehingga teks dapat diolah lebih lanjut.

2. Pembobotan Data

Pembobotan data dilakukan dengan TF-IDF (*Term Frequency-Inverse Document Frequency*) yaitu skema pembobotan yang digunakan untuk menghilangkan istilah dengan bobot lebih rendah dari dokumen dan membantu meningkatkan efektivitas pengambilan. frekuensi dokumen terbalik (IDF) adalah statistik numerik yang memberi tahu kita seberapa penting suatu kata bagi suatu dokumen dalam koleksi atau korpus.

3. K-Means Clustering Data

Metode analisis lain yaitu dengan K-Means *Clustering* data yaitu mengelompokkan data komentar dari user TikTok dengan cara mengelompokkan sebuah unit objek data ke dalam kumpulan bagian yang dikenal sebagai *cluster* (kelas), yang memberi peluang bagi data dalam satu *cluster* memiliki tingkat kemiripan yang lebih tinggi dan data di antara *cluster* memiliki tingkat kemiripan lebih rendah.

4. Elbow Method

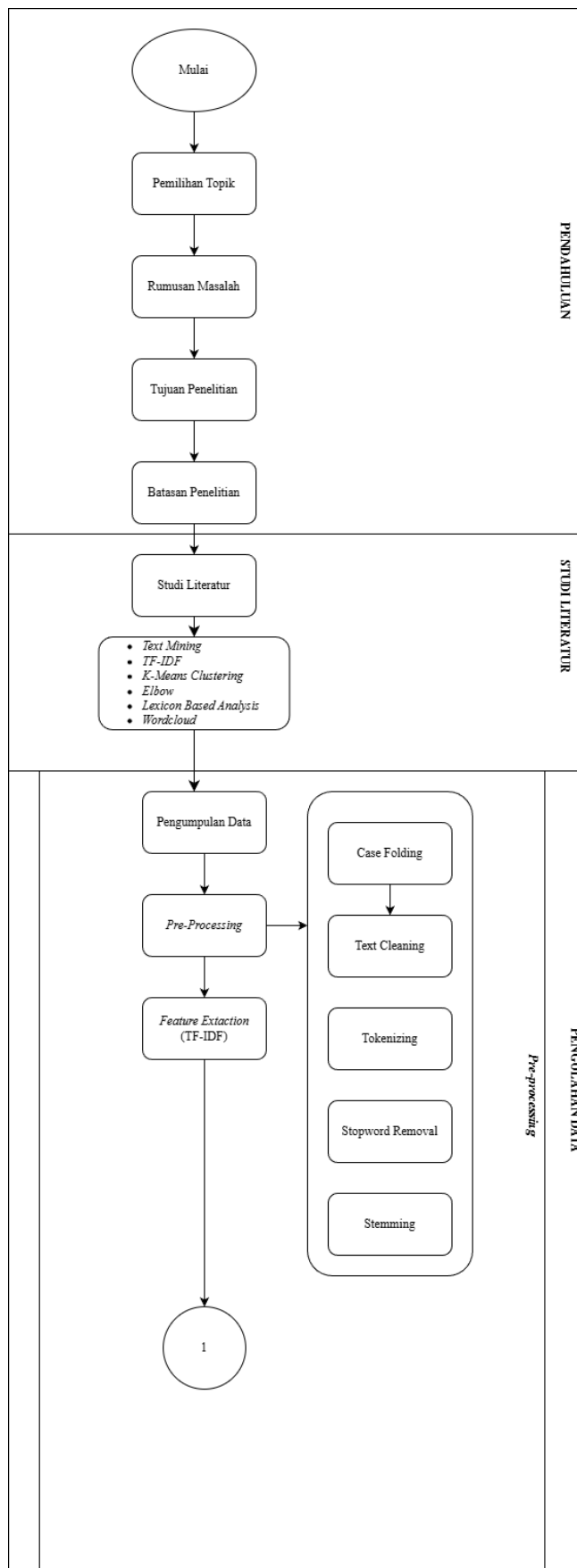
Metode Elbow ini merupakan metode yang digunakan untuk menentukan jumlah *cluster* (kelas) terbaik. Metode ini dilakukan dengan cara melihat presentase setiap *cluster* yang akan membentuk siku pada suatu titik tertentu.

5. Analisis Sentimen *Lexicon Based*

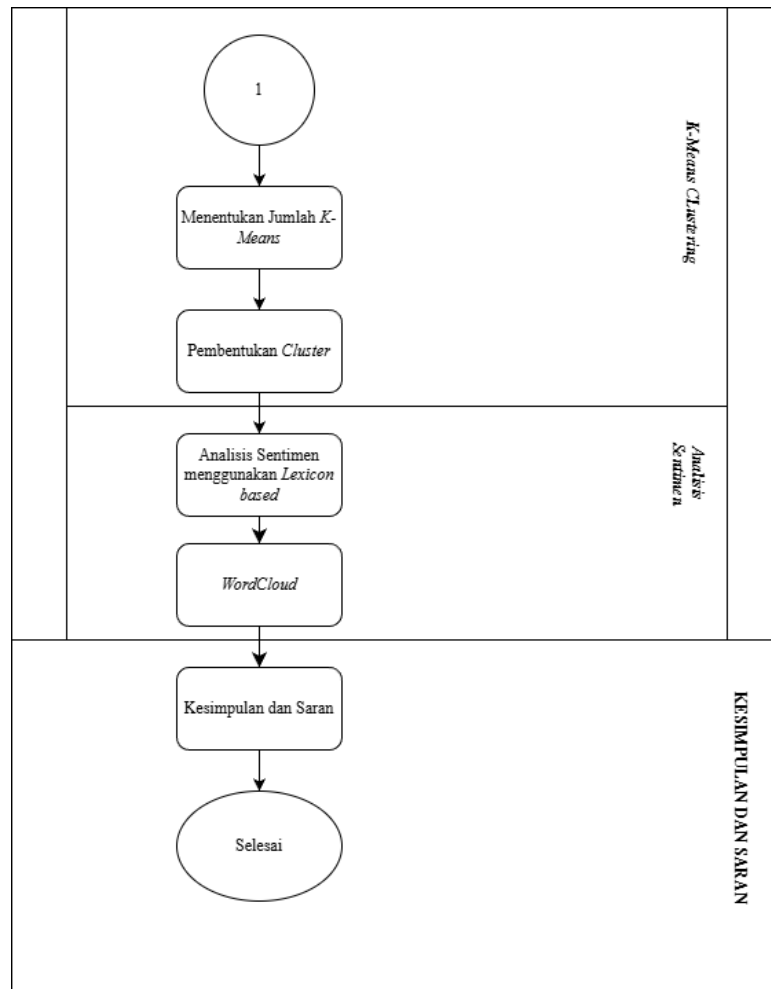
Metode selanjutnya, terdapat analisis sentimen atau biasa disebut juga *opinion mining* merupakan metode yang menganalisis opini, sentimen, penilaian, sikap, dan emosi manusia terhadap suatu entitas dan atribut yang diekspresikan dalam teks tertulis. Penerapan analisis sentimen ini karena Analisis sentimen pada penelitian ini menggunakan algoritma *Lexicon based* yaitu pendekatan analisis berdasarkan *Lexicon* untuk klasifikasi ekspresi sentimen yang meliputi kata sentimen, frasa, idiom, dan aturan komposisi.

3.7 Alur Penelitian

Alur penelitian pada penelitian ini digunakan sebagai acuan untuk proses penelitian dari awal penentuan topik hingga hasil akhir analisis kesimpulan penelitian. Tahapan ini disusun dengan baik agar segala kegiatan yang dilakukan dapat dilaksanakan sesuai rencana penelitian dan tetap pada standar masing-masing penelitian. Diharapkan dari adanya alur penelitian ini, peneliti dapat dengan terstruktur menyelesaikan penelitian dengan seksama dari awal hingga akhir penelitian. Berikut merupakan serangkaian tahap diagram alir untuk penelitian ini.



Gambar 3. 1 Alur Penelitian (1)



Gambar 3. 2 Alur Penelitian (2)

Gambar di atas merupakan alur penelitian ini, Gambar 3. 1 dan Gambar 3. 2 memiliki penjelasan alur sebagai berikut:

1. Mulai
2. Pemilihan Topik (identifikasi masalah)

Pemilihan topik dimulai dari identifikasi masalah yang ada. Permasalahan yang akan diangkat pada penelitian ini terkait Kebijakan Sertifikasi Halal di Indonesia. Untuk menentukan topik tersebut perlu diidentifikasi permasalahan yang ada melihat situasi dan kondisi dari dampak adanya kebijakan sertifikasi halal ini bagi masyarakat Indonesia melalui opini yang disampaikan oleh publik, dampaknya, maupun peningkatan yang terjadi dari ditetapkan kebijakan tersebut.

3. Merumuskan Masalah

Langkah selanjutnya yaitu merumuskan masalah dengan setelah identifikasi permasalahan tersebut, lalu melakukan pencarian akar permasalahan dari topik yang diangkat di penelitian ini. Untuk nantinya *dilanjutkan* ke tahap tujuan hingga manfaat di penelitian ini.

4. Menentukan Tujuan, Batasan Penelitian, Manfaat Penelitian

Pada tujuan berisi jawaban atau hal yang akan dicapai dari rumusan masalah yang sudah dibuat. Dan sebagai landasan untuk penelitian ini. Kemudian penelitian ini memiliki batasan penelitian yang berfungsi agar peneliti melakukan penelitian ini fokus hanya terhadap penelitian yang sedang dibahas sehingga tidak melebar pada topik lain, seperti jumlah sampel sebanyak 1100 komentar. Komentar ini dimulai sejak tanggal 2 Februari 2024 hingga 29 November 2024, dengan komentar Bahasa Indonesia. Lalu terakhir terdapat manfaat yang diperoleh jika penelitian ini dilakukan.

5. Studi Literatur

Pada alur ini, dilakukan pencarian dan telaah studi literatur sebagai landasan teori dan memperkuat data dan pengolahan data yang akan dilakukan. Serta mencari informasi yang tepat dengan topik penelitian saat ini. Dan sebagai kajian untuk penelitian-penelitian terdahulu yang relevan dengan topik penelitian yang akan dilakukan.

6. Pengumpulan Data

Pengumpulan data berfungsi untuk memperoleh data yang akan diolah dan dianalisis. Pengumpulan data pada penelitian ini menggunakan *Apify* dengan *TikTok Comment Scraper*, dengan mengambil komentar yang berasal dari komentar pada konten akun TikTok yaitu user @fathianhafiz <https://vt.tiktok.com/ZSPFUuwH/> dan @medeoind.media <https://vt.tiktok.com/ZSPFUTwE5/>. Hasil dari data yang diambil tersebut berupa komentar dari dua konten tersebut yang diambil sebanyak 1100 komentar, kemudian nantinya akan diolah pada proses *pre-processing*.

7. Pre-Processing Data

Pada tahap ini dilakukan proses pembersihan data dengan menghapus data-data yang tidak diperlukan atau tidak relevan. Setelah data dibersihkan akan menjadi akurat dan terstruktur. Berikut merupakan tahapan dari *pre-processing*:

a. *Case Folding*

Proses ini *Case Folding* ini dilakukan saat teks yang terdiri dari berbagai huruf besar dan kecil distandarisasi. Contoh kata sebelum dan sesudah dilakukan *case folding* yaitu 'Tempat' menjadi 'tempat,' dan 'Tahun' menjadi 'tahun.'

b. *Text Cleaning*

Pada proses ini dilakukan untuk membersihkan teks dari noise-noise dan outlier yang tidak perlu, menghapus kata-kata atau kalimat yang tidak relevan dengan topik. Pada proses ini dilakukan dengan memperbaiki data untuk normalisasi data, dengan menggunakan bahasa baku misalnya 'yg' menjadi 'yang.'

c. *Tokenizing*

Proses ini dilakukan untuk memisahkan suatu kalimat menjadi berbagai pecahan kata. Misalnya [aku tidak setuju dengan aturan ini] menjadi [aku', tidak', setuju', dengan', aturan', ini.']

d. *Stopword Removal*

Pada tahap ini, kata-kata yang dianggap tidak bermakna dan sering muncul dalam jumlah yang besar dalam data akan dihilangkan. Contoh kata yang dihilangkan yaitu 'dan,' 'atau,' 'yang.'

e. *Stemming*

Stemming digunakan untuk mengubah kata menjadi bentuk akarnya. Contoh dalam penelitian. Menggunakan perpustakaan sastrawi yang memiliki kamus akar kata Indonesia, contohnya proses stemming dari kata 'menyenangkan' menjadi kata dasar 'senang.'

8. *Feature Extraction*

Pada tahap ini dilakukan dengan mengubah atau mengekstraksi data mentah menjadi fitur yang penting yang memudahkan algoritma untuk mengambil keputusan atau analisis data.

9. *Clustering*

Data yang sudah dilakukan proses pembersihan atau pre-processing, kemudian data akan dilakukan proses pengelompokan data pada masing-masing *cluster* menggunakan metode *K-Means* sebagai berikut:

- a. Menentukan jumlah *cluster* (k) menggunakan metode Elbow.
- b. Mengalokasikan objek ke dalam *cluster* secara acak.

- c. Menghitung *centroid* sampel yang ada pada masing-masing *cluster*, menggunakan *euclidean distance*.
- d. Mengalokasikan objek ke dalam masing-masing centroid terdekat.
- e. Kemudian, kembali ke langkah 3 jika terdapat objek yang masih ada perubahan nilai *centroid*, atau di atas nilai *threshold* yang ditentukan, ataupun jika perubahan nilai *objective function* yang digunakan di atas nilai *threshold*.

10. Analisis menggunakan *Lexicon Based*

Setelah melakukan proses *clustering* dan mendapat kelas masing-masing, kemudian dilakukan analisis sentimen menggunakan analisis berdasarkan *lexicon*. Analisis sentimen menggunakan *lexicon* ini untuk klasifikasi ekspresi sentimen yang meliputi kata sentimen, frasa, idiom, dan aturan komposisi. Dan hasil inverse perkalian digunakan untuk memberikan bobot yang lebih rendah pada ekspresi sentimen yang jaraknya lebih jauh dari aspek a_1 . Jika skor akhir bernilai positif, maka opini terhadap aspek a_1 bersifat positif, namun jika bernilai negatif maka sentimen akan negatif, jika tidak maka, bersifat netral.

11. *WordCloud*

Pada tahap *WordCloud* ini merupakan visualisasi kata-kata dari sebuah teks, yang paling sering muncul saat ditampilkan, dan kata-kata yang memiliki huruf lebih besar merupakan yang sering muncul.

12. Kesimpulan dan Saran

Jika tahap analisis dan pengolahan data sudah dilakukan dan mendapatkan hasil, maka langkah selanjutnya adalah membuat kesimpulan dari hasil tersebut. Kemudian, membuat saran bagi penelitian selanjutnya agar dapat menjadi pengembangan untuk penelitian mereka.

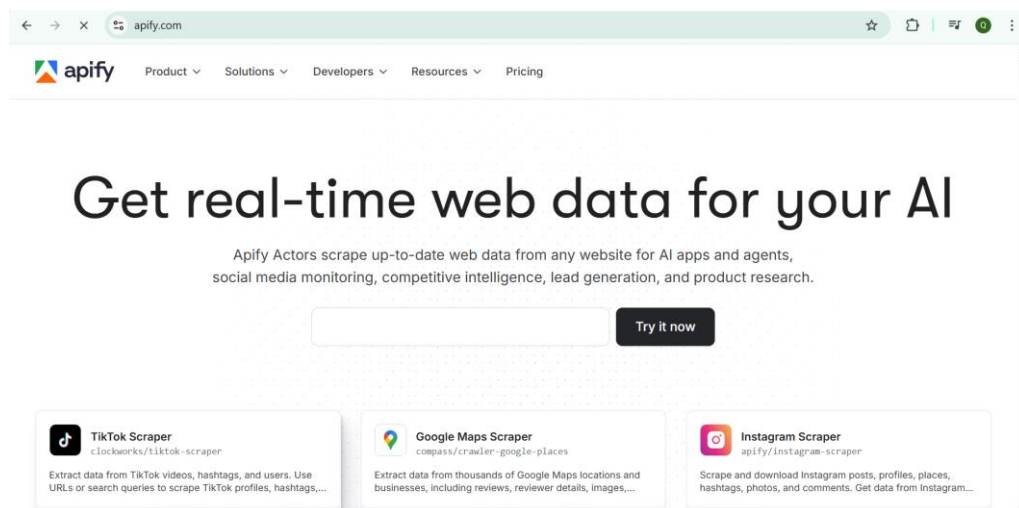
13. Selesai

BAB IV

PENGUMPULAN DAN PENGOLAHAN DATA

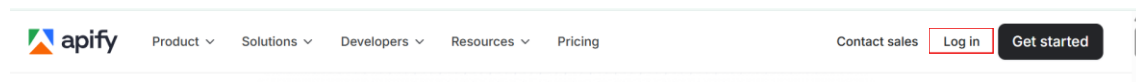
4.1 Pengumpulan Data

Pada pengumpulan data penelitian ini dilakukan dengan menggunakan data yang diperoleh dari proses *scrapping* data (mengeksktrak data) dari platform di internet yaitu *website* Apify. Proses *scrapping* ini mengambil data dari platform TikTok dengan hanya mengambil data pada bagian fitur komentar saja. Scrapping data pada platform Apify ini memiliki berbagai macam tools yang dapat dipilih untuk melakukan proses ekstrak data yang kemudian bisa mengubah nya menjadi file jenis tertentu.



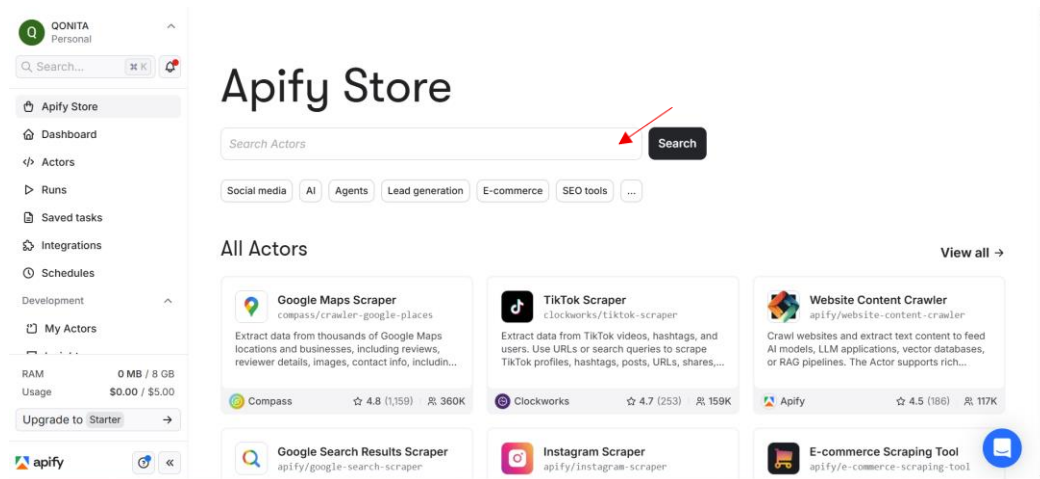
Gambar 4. 1 *Interface* Platform Apify

Gambar 4. 1 merupakan *Interface* dari platform Apify, yang memiliki banyaknya fitur ekstraksi data, salah satunya yang digunakan dalam penelitian ini adalah *TikTok Comment Scraper*. Untuk dapat menemukan fitur tersebut, langkah utama yang dilakukan adalah dengan melakukan *Log in* seperti pada Gambar 4. 2 di bawah ini.

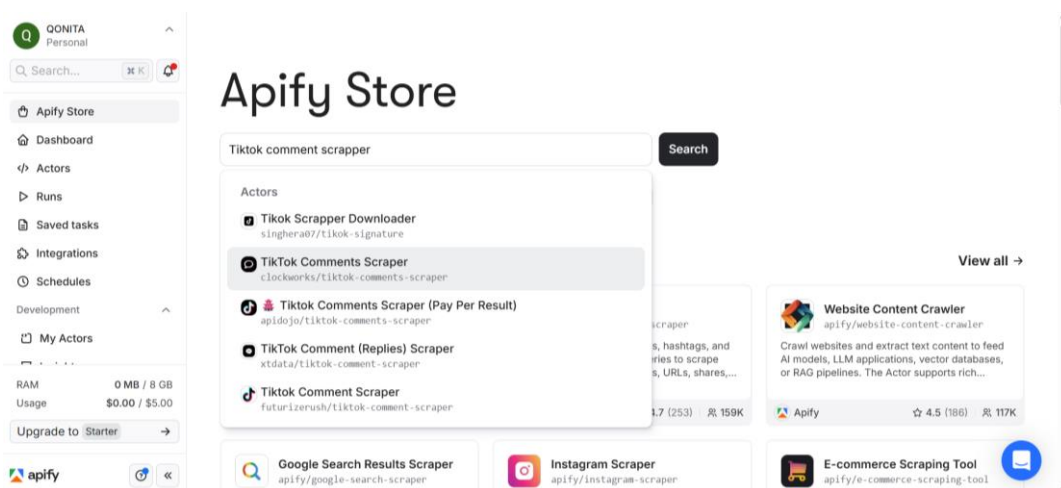


Gambar 4. 2 *Log In* di Apify

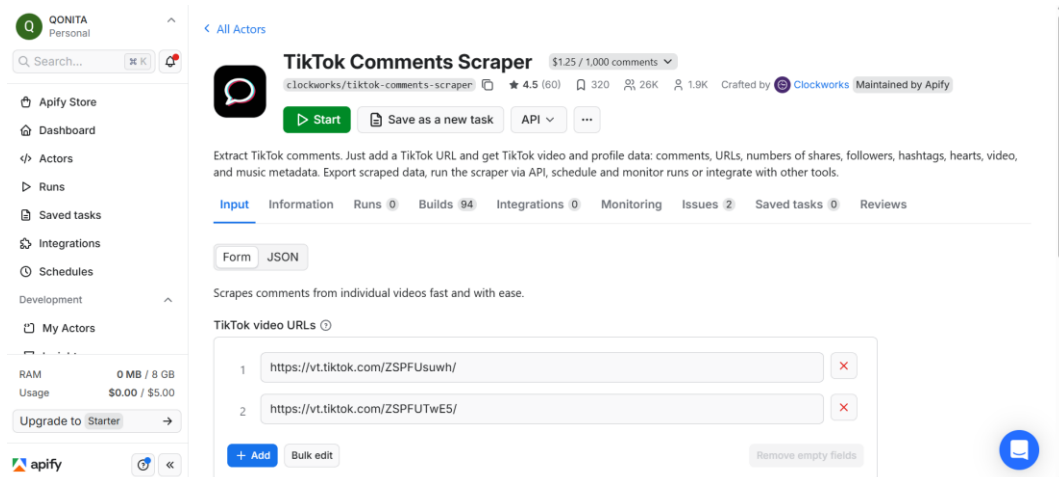
Setelah melakukan proses ‘Log In’ maka akan diarahkan ke tampilan pada Gambar 4. 3. Setelah tampilan terbuka, kemudian akan muncul beberapa tools, salah satunya yaitu laman pada ‘Apify Store.’ Pada bagian ini, yang dilakukan adalah menuliskan fitur *actors* yang akan berfungsi untuk proses scrapping pada tools ‘Search Actors’ seperti yang ditunjuk oleh panah merah.

Gambar 4. 3 *search actors*

Pada tools ini, menuliskan *actors* yang dicari, karena penelitian ini mengambil data dari komentar atau opini di suatu konten TikTok maka *actors* yang dipakai yaitu TikTok Comment *Scraper* seperti Gambar 4. 4.

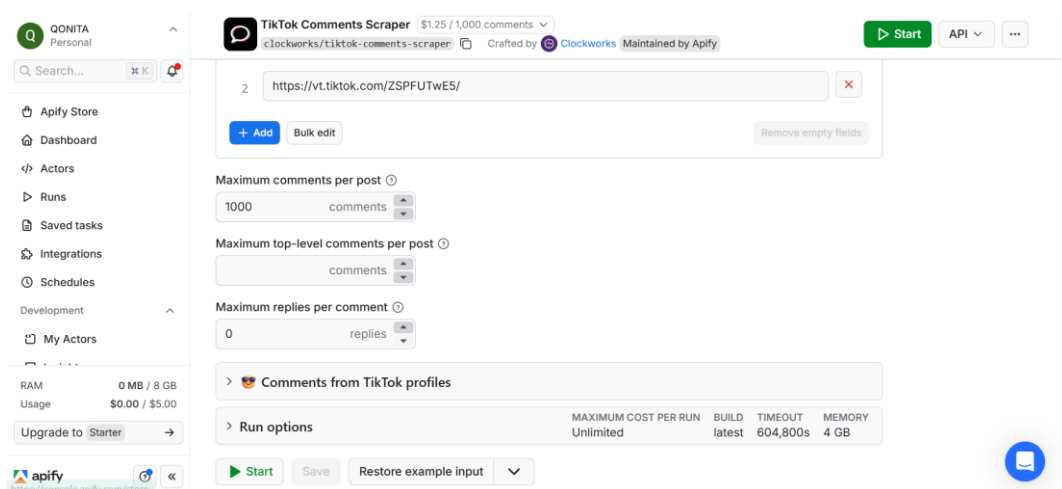
Gambar 4. 4 menuliskan TikTok Comment *Scraper*

Dari proses search tersebut, akan muncul tampilan Gambar 4. 5 dan Gambar 4. 6 di bawah ini, yang harus dilakukan yaitu memasukkan input. Input yang digunakan adalah TikTok video URLs atau Link dari video konten TikTok. Pada penelitian ini menggunakan dua URLs vide konten, untuk memasukkan link tersebut terdapat tombol +Add berwarna biru untuk memasukkan URLs dua video yang digunakan pada penelotian ini, seperti tampilan yang sudah ada di Gambar 4. 5 tersebut yaitu link URLs dari user @fathianhafiz <https://vt.tiktok.com/ZSPFUsuwh/> dan @medeoind.media <https://vt.tiktok.com/ZSPFUTwE5/>.



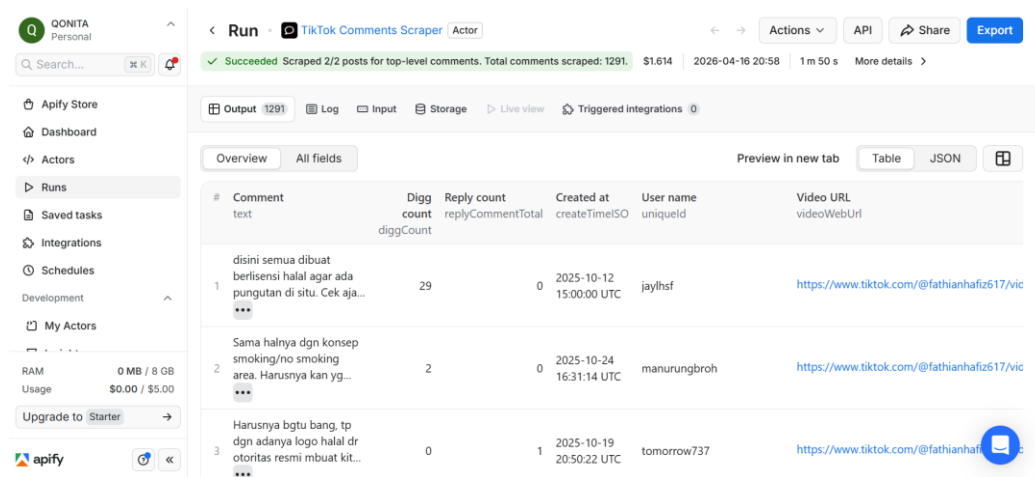
Gambar 4. 5 TikTok Comment *Scraper* actors

Kemudian di bagian bawah setelah memasukkan URLs video, akan ada kolom informasi yang diperlukan yaitu dengan memasukkan input Maximum komentar yang diinginkan untuk diambil sebagai data penelitian pada Gambar 4. 6. Dalam penelitian ini, maximum komentar yang diambil yaitu 1000 komentar per postingan video. lalu, setelah menuliskan maximum komentar yang diinginkan, langkah selanjutnya adalah melakukan klik ‘Start’ untuk melakukan proses Running.



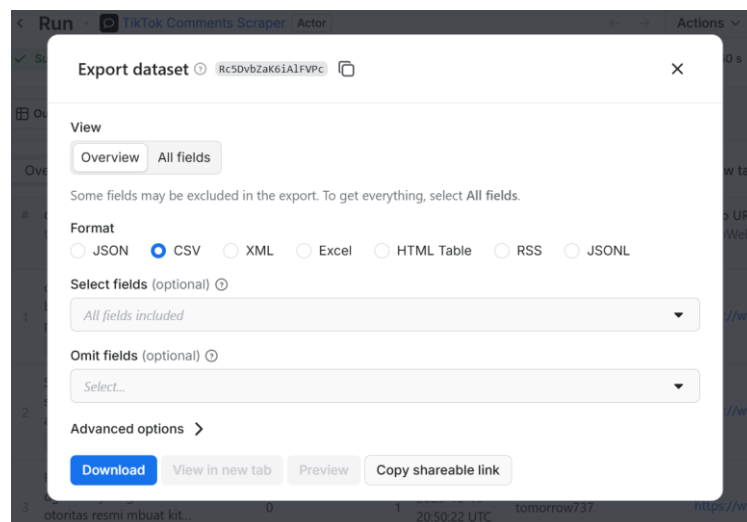
Gambar 4. 6 Proses *Running*

Proses Running akan yang dinyatakan ‘Succeeded’ seperti Gambar 4. 7 di bawah.



Gambar 4. 7 *Running* selesai dan *Export* file

lalu hasil *Scraper* dinyatakan selesai, kemudian dapat diekstrak menjadi file csv. Melalui proses klik ‘Export’ pada bagian tombol berwarna biru, dan memilih format CSV, lalu klik tombol ‘Download,’ maka file akan tersimpan.



Gambar 4. 8 Download File csv

Hasil dari data yang diambil tersebut berupa komentar dari dua konten sebanyak 1100 komentar terambil dari tanggal 2 Februari 2024 hingga 29 November 2024, dengan hasil file csv pada gambar 11 berikut ini.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	text	digCount	replyComr	createTim	uniqueId	videoWebU	cid	avatarThumb											
2	disini semi	29	0	2025-10-1	jaythst	https://ww	6.99E+18	7.56E+18	https://p16-common-sign.tiktokcdn-us.com/tos-alisg-avt-0068/ee28c37d76afc56d0c0adbab5e99606d-tpv-tiktok-cropc										
3	Sama haln	2	0	2025-10-2	manunungl	https://ww	7.04E+18	7.56E+18	https://p16-common-sign.tiktokcdn-us.com/tos-maliva-avt-0068/4c0425ee6bc7e449e1d77db3c6f384-tpv-tiktok-crop										
4	Harussyay i	0	1	2025-10-1	tomorrow	https://ww	6.82E+18	7.56E+18	https://p16-common-sign.tiktokcdn-us.com/tos-maliva-avt-0068/cc89ea095b0d359ff28454eacaf6164-tpv-tiktok-crop										
5	simpler aja	3	0	2025-10-1	fnndf	https://ww	7.29E+18	7.56E+18	https://p16-common-sign.tiktokcdn-us.com/musically-maliva-obj/1594805258216454-tpv-tiktok-cropcenter:100:100.jp										
6	gua gk seti	8	5	2025-10-1	bluesky.or	https://ww	6.81E+18	7.56E+18	https://p16-common-sign.tiktokcdn-us.com/tos-useast2a-avt-0068/gjsor/7f23400f5a9b9a3d09a35079337691f5-tpv-tiktok										
7	bener juga	7	0	2025-10-1	coba.dulu	https://ww	7.46E+18	7.56E+18	https://p19-common-sign.tiktokcdn-us.com/tos-alisg-avt-0068/ec3a8faad12788d7c5067fed8b8dc03-tpv-tiktok-cropce										
8	Kali ini gw i	3	1	2025-11-0	andfaiz96	https://ww	7.17E+18	7.57E+18	https://p16-common-sign.tiktokcdn-us.com/tos-useast2a-avt-0068-euttp/7389678096713051169-tpv-tiktok-cropcenter										
9	katau kata	1	2	2025-10-1	pdtkubdm1	https://ww	7.44E+18	7.57E+18	https://p16-common-sign.tiktokcdn-us.com/tos-alisg-avt-0068/19bd7b24b9ab60ec39f300b5807e7b-tpv-tiktok-cropce										
10	lh buat pro	2	0	2025-10-1	cb150r.or	https://ww	7.05E+18	7.56E+18	https://p16-common-sign.tiktokcdn-us.com/tos-alisg-avt-0068/2a34290c0d54e86ecc3949d68ce8905-tpv-tiktok-cropc										
11	gak gitu loj	11	6	2025-10-1	buzzlighty	https://ww	7.13E+18	7.56E+18	https://p16-common-sign.tiktokcdn-us.com/tos-alisg-avt-0068/55ef928b5025eaa8aaf2a25821d46d1-tpv-tiktok-cropi										
12	Kemaren u	1	2	2025-10-1	uncov3r	https://ww	7.06E+18	7.56E+18	https://p16-common-sign.tiktokcdn-us.com/tos-alisg-avt-0068/8baccbe2a023c9c156cc1ad6fb3d41651-tpv-tiktok-cropc										
13	naman	1	0	2025-12-0	elywydiaw	https://ww	7.29E+18	7.56E+18	https://p16-common-sign.tiktokcdn-us.com/tos-alisg-avt-0068/0ce8a4bd04221eb16abb5a1a9a6db1-tpv-tiktok-cropc										
14	berapa baw	2	3	2025-10-2	user25036	https://ww	7.5E+18	7.56E+18	https://p16-common-sign.tiktokcdn-us.com/tos-alisg-avt-0068/a17d95f347e4557c51fc64f8690c4c0-tpv-tiktok-cropce										
15	ya kan bua	1	1	2025-10-1	holongsun	https://ww	6.57E+18	7.56E+18	https://p16-common-sign.tiktokcdn-us.com/tos-maliva-avt-0068/6f703c6dfa0154bc06ba4f7fcd2b5d6-tpv-tiktok-cropci										
16	ya emang i	0	1	2025-10-1	arby_wp1	https://ww	7.42E+18	7.56E+18	https://p16-common-sign.tiktokcdn-us.com/tos-maliva-avt-0068/edd81da7eaa111422aaa976363acbd7-tpv-tiktok-croq										

Gambar 4. 9 File format csv yang sudah di download

4.2 Pengolahan Data

Pada bagian pengolahan data ini dilakukan setelah semua data terkumpul sesuai langkah yang sudah dilakukan di bagian 4.1. pengolahan data dilakukan menggunakan file csv yang sudah *scrapping*, kemudian langkah selanjutnya masuk ke dalam tahap pre-processing menggunakan *tools Python* dan *Google Colab*. Berikut ini merupakan langkah dilakukannya tahap pre-processing data.

4.2.1 Pre-Processing Data

Dalam tahap ini, dilakukan proses pembersihan data dari data tidak relevan atau informasi yang tidak relevan, seperti gangguan pada teks sebelum masuk ke dalam tahap pengolahan data, dengan variabel penting yang diperlukan yaitu 'UniqueId' sebagai ID pengguna TikTok, 'CreateTimeISO' untuk mengetahui waktu komentar dibuat, lalu 'Text' sebagai komentar atau opini pengguna. Urutan tahap pembersihan data ini dilakukan melalui tahapan berikut ini.

a. Case folding

Tahap *case folding* ini dilakukan untuk menyamaratakan atau standarisasi berbagai huruf besar (kapital) menjadi teks dengan huruf kecil dalam suatu teks. Berikut contoh teks di penelitian ini tertera pada Tabel 4. 1, pada huruf atau kata bertanda merah tersebut masih dalam huruf kapital, untuk itu perlu distandarisasi menjadi huruf kecil (*lowercase*).

Tabel 4. 1 *Case Folding* (sebelum dan Sesudah)

<i>Input</i>	<i>Output</i>
Berarti anda TIDAK ada Toleransi antar umat Beragama . Indonesia ini sangat beragam (beda suku, adat, Ras , Bahasa , AGAMA , dll) Indonesia BUKAN negara AGAMA TERTENTU . Indonesia itu Bhineka tunggal Ika .	berarti anda tidak ada toleransi antar umat beragama. indonesia ini sangat beragam (beda suku, adat, ras, bahasa, agama, dll) indonesia bukan negara agama tertentu. indonesia itu bhineka tunggal ika.
Alhamdulillah... aku udah daftar bg. cuma jualan es teh. tapi udah ada sertifikat nya, kemarin gratis bg.	alhamdulillah... aku udah daftar bg. cuma jualan es teh. tapi udah ada sertifikat nya, kemarin gratis bg.
ada cap HALAL , tapi ada tulisan MENGANDUNG BABI wkwk	ada cap halal, tapi ada tulisan mengandung babi wkwk
Tp sekarang masih gratis bang	tp sekarang masih gratis bang

b. Text Cleaning

Pada tahap ini, setelah dilakukan *case folding*, maka data akan dibersihkan dari gangguan atau noise yang tidak diperlukan. Penelitian hanya menggunakan komentar yang dibuat hanya pada tanggal 2 Februari 2024 hingga 29 November 2024. Sehingga data yang diolah memiliki batasan hanya dari bulan Februari 2024 hingga November 2024, dan hasil cleaning pada Gambar 4. 10 *Text Cleaning* berikut.

	text	diggCount	replyCommentTotal	createTimeISO	uniqueId	cleaning
0	Yang dagang babi keliling gimana? 🐷	13	5	2024-02-02T14:46:39.000Z	callmemoriz	yang dagang babi keliling gimana
1	Beli bahannya mesti sama supplier yg ditunjuk ...	1	0	2024-02-02T16:25:59.000Z	eddylin168	beli bahannya mesti sama supplier yg ditunjuk ...
2	buat beli rubikon lgi	313	7	2024-02-02T05:05:22.000Z	nurshina24	buat beli rubikon lgi

Gambar 4. 10 *Text Cleaning*

c. *Tokenizing*

Pada proses tokenizing ini digunakan untuk memisahkan kata dalam suatu kalimat menjadi pecahan kata tertentu sesuai yang ada di dalam kalimat. Hasil pemisahan menjadi pecahan kata ini tertera dalam contoh Tabel 4. 2.

Tabel 4. 2 *Tokenizing*

<i>Input</i>	<i>Output</i>
alhamdulillah... aku udah daftar bg. cuma jualan es teh. tapi udah ada sertifikat nya, kemarin gratis bg.	['alhamdulillah', 'aku', 'udah', 'daftar', 'bg', 'cuma', 'jualan', 'es', 'teh', 'tapi', 'udah', 'ada', 'sertifikat', 'nya', 'kemarin', 'gratis', 'bg']
ada cap halal , tapi ada tulisan mengandung babi wkwk	['ada', 'cap', 'halal', 'tapi', 'ada', 'tulisan', 'mengandung', 'babi', 'wkwk']
tp sekarang masih gratis bang	['tp', 'sekarang', 'masih', 'gratis', 'bang']

d. *Stopword Removal*

Proses stopword removal ini berfungsi dalam pengolahan data untuk menghentikan penggunaan kata yang tidak bermakna atau sering muncul dalam jumlah yang besar kemudian akan dihilangkan. Contoh kata yang dihentikan pada proses ini yaitu kata umum seperti "supaya", "jadi", "karena", kata filler "aja", "doang", "banget", " dan slang seperti "yg", "dg", "dgn", serta sosmed noise seperti "wkwk", "wk", "haha" pada Tabel 4. 3

Tabel 4. 3 *Stopword Removal*

<i>Input</i>	<i>Output</i>
tp sekarang masih gratis bang	sekarang masih gratis bang
270 juta warga kuota gratis 1jt doang	270 juta warga kuota gratis 1jt
ada cap halal tapi ada tulisan mengandung babi wkwk	ada cap halal tapi ada tulisan mengandung babi

e. *Stemming*

Pada proses *stemming*, berfungsi mengubah kata menjadi bentuk aslinya atau akarnya, dalam penelitian ini menggunakan perpustakaan sastrawi yang memiliki

kamus akar kata Indonesia. Dalam Tabel 4. 4 berikut meruoakan hasil dari proses stemming pada penelitian ini.

Tabel 4. 4 *Stemming*

<i>Input</i>	<i>Output</i>
apakah halal itu bisa jadi jaminan qt di selamatkan	['apakah', 'halal', 'bisa', 'jamin', 'kita', 'selamat']
adek gw kerjanya kaya gtu , bikin sertifikat kaya gtu , tapi gratis	['adik', 'aku', 'kerja', 'seperti', 'gitu', 'bikin', 'sertifikat', 'seperti', 'gitu', 'tapi', 'gratis']
pajak berkedok sertifikat	['pajak', 'kedok', 'sertifikat']

f. *Feature Extraction (TF-IDF)*

Selanjutnya setelah tahap *stemming* data, dilakukan tahap penambahan fitur *Term Frequency – Inverse Document Frequency* untuk mengubah atau ekstraksi data mentah menjadi fitur penting yang memudahkan algoritma dalam mengambil keputusan atau analisis data nantinya dalam proses *clustering* dan analisis sentimen.

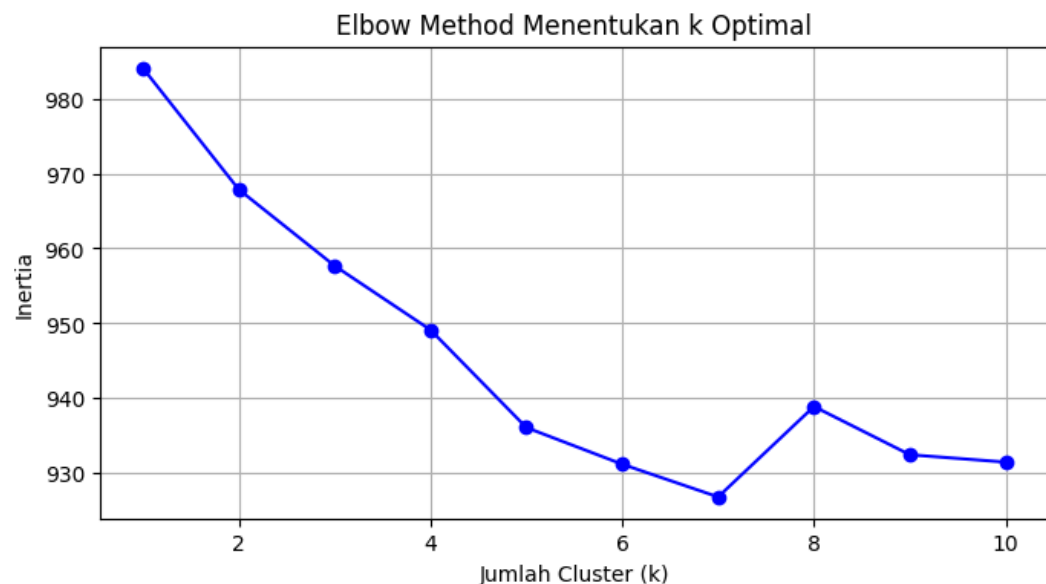
g. *Clustering menggunakan Algoritma K-Means*

Proses *clustering* dilakukan menggunakan algoritma *K-Means*, dalam tahap *clustering* ini terdapat proses menambahkan pembobotan kata menggunakan fitur TF-IDF, yang bertujuan untuk memberikan bobot pada setiap kata dalam teks komentar yang berguna agar memudahkan identifikasi kata yang sesuai dan relevan dengan masalah terkait kebijakan sertifikasi halal ini. berikut ini Gambar 4. 11 menunjukkan pembobotan yang selanjutnya digunakan untuk menentukan kelas dalam *clustering*.

...	Coords	Values	(3, 1114)	0.4149828707926043
	(0, 102)	0.3497854556854193	:	:
	(0, 358)	0.4482193415945529	...	(1095, 1711) 0.26915964846608414
	(0, 489)	0.40835475768405005		(1096, 228) 0.3911238545133149
	(0, 734)	0.7141400063631008		(1096, 1175) 0.47167340555728104
	(1, 64)	0.3665457234770267		(1096, 1338) 0.2780697588949866
	(1, 114)	0.3449617176711802		(1096, 1386) 0.5497727380794266
	(1, 179)	0.2945586387594701		(1096, 1513) 0.43938791426811713
	(1, 359)	0.2727120967511694		(1096, 1588) 0.2278414698172658
	(1, 962)	0.3513375565696259		(1097, 17) 0.31961697891210583
	(1, 1344)	0.3030162627864047		(1097, 140) 0.3476549916090592
	(1, 1509)	0.41781952963768615		(1097, 534) 0.11127247050735171
	(1, 1625)	0.4423560478209708		(1097, 579) 0.37614859099016085
	(2, 179)	0.4768772863832891		(1097, 719) 0.24618143751186863
	(2, 271)	0.4133244188060843		(1097, 829) 0.31961697891210583
	(2, 830)	0.3797260277118115		(1097, 855) 0.27246714021166196
	(2, 1326)	0.6764311660683086		(1097, 1014) 0.23854414613936706
	(3, 6)	0.18023013846262792		(1097, 1380) 0.25275939649099843
	(3, 114)	0.32361534247785584		(1097, 1449) 0.31961697891210583
	(3, 534)	0.1227604472205349		(1097, 1588) 0.14723998994715792
	(3, 547)	0.2763312271445149		(1097, 1685) 0.37614859099016085
	(3, 554)	0.1869734604659448		(1098, 30) 0.44516070051822754
	(3, 672)	0.19012815187069412		(1098, 82) 0.48072728472566434
	(3, 710)	0.26317188724568286		(1098, 99) 0.41243637566128344
	(3, 766)	0.3919646826043525		(1098, 102) 0.22564929387316052
	(3, 1114)	0.4149828707926043		(1098, 534) 0.3011204788754131
:	:	:		(1098, 1571) 0.5089580708095526

Gambar 4. 11 Bobot Frekuensi TF-IDF

Selanjutnya, proses *clustering* dilakukan menggunakan algoritma *K-Means*, juga menerapkan metode Elbow untuk memperoleh hasil kelas yang optimal, seperti Gambar 4. 12 berikut, untuk menentukan kelas yang optimal dapat dilihat dari titik siku penurunan tajam dimana nilai inertia akan mulai melambat.



Gambar 4. 12 Jumlah Cluster

Dari Gambar 4. 12 tersebut, dapat dilihat bahwa titik siku dalam grafik tersebut mengalami penurunan tajam di angka 5 yang berarti penurunan ini dinilai cukup signifikan, dan penurunan setelahnya lebih landai dianggap tidak memberi

peningkatan yang berarti dalam pemodelan data. Jadi kesimpulan kelas (K) optimal dalam penelitian ini menggunakan jumlah *cluster* K = 5.

h. Pemeriksaan data NaN (*null*)

Dalam proses pemeriksaan data NaN ini, biasanya setelah proses cleaning itu terdapat noise atau gangguan seperti baris kosong atau duplikat yang ada pada data. Setelah memeriksa data NaN yang ada, kemudian menghapus data tersebut. Maka dalam penelitian ini semula data berjumlah 1030 menjadi 961 data, jadi yang sudah terhapus terdapat 69 data *null* terhapus. Dalam cek jumlah data NaN maka menggunakan bahasa pemrograman berikut.

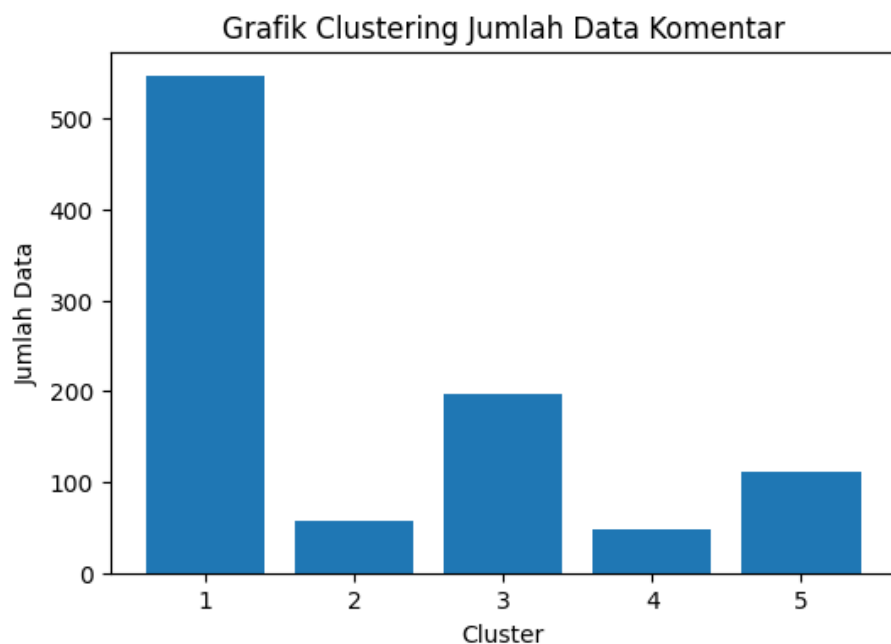
```
jumlah_nan = df['text_final'].isna().sum()
print("Jumlah data NaN:", jumlah_nan)
```

... Jumlah data NaN: 0

Gambar 4. 13 Pemeriksaan Data Nan (null)

i. Pembentukan *Cluster*

Dalam pembentukan *cluster*, diperoleh hasil *cluster* K = 5 setelah diterapkan metode Elbow untuk data komentar. Maka dari diperolehnya jumlah *cluster* yang optimal sebanyak 5, menggunakan bahasa pemrograman python dan proses *clustering* menggunakan K-Means, diperoleh jumlah data komentar yang masuk dalam setiap kelas sebagai berikut sesuai.



Gambar 4. 14 Grafik Jumlah Clustering

```

cluster
1      546
2       58
3      197
4       48
5      112
Name: count, dtype: int64

```

Gambar 4. 15 Jumlah masing-masing *cluster*

Pada **Error! Reference source not found.** merupakan jumlah pada tiap masing-masing *cluster* yang terbentuk. Pada *cluster* 1 terbentuk *cluster* dengan jumlah 546 data, dan *cluster* 2 sebanyak 58 data, *cluster* 3 sebanyak 197 data, *cluster* 4 sebanyak 48 data, Terakhir *cluster* 5 sebanak 112 data. Dengan total keseluruhan data berjumlah 961 komentar data konsumen.

Selanjutnya, setelah proses dan hasil dari *clustering* ini didapatkan, langkah selanjutnya adalah identifikasi ciri khas setiap *cluster* menggunakan visualisasi *WordCloud*. *WorldCloud* membantu secara visual menggambarkan frekuensi dan signifikansi relatif dari kata-kata dalam teks, memudahkan pemahaman tema atau kata kunci dan fokus utama dengan cepat (Agusia et al., 2024). Visualisasi *WordCloud* ini berfungsi untuk menunjukkan kata, informasi, atau topik yang paling sering muncul atau dibicarakan oleh konsumen dalam data komentar TikTok ini.

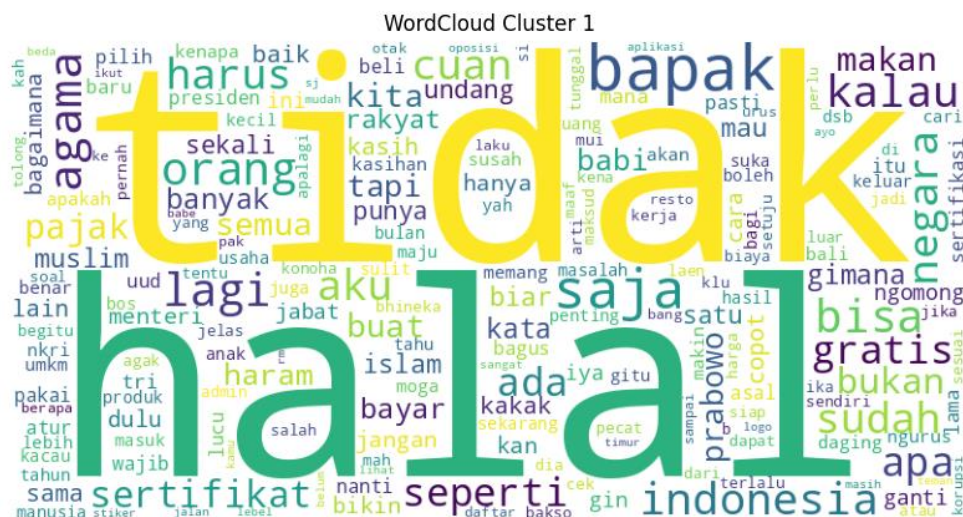
4.3 Identifikasi Ciri Khas setiap *Cluster*

Proses pengolahan data dilakukan dengan *pre-processing* kemudian *clustering* menggunakan metode *K-Means*, dan bantuan algoritma *Elbow* untuk menentukan kelas paling optimal. Setelah itu, tahapan selanjutnya hasil *clustering* ini kemudian dilakukan visualisasi menggunakan *WordCloud* untuk mengetahui tampilan topik, kata, atau informasi yang paling sering dibahas dalam data komentar konsumen. Lalu, dilakukan pembobotan dalam identifikasi ciri khas setiap *cluster* dengan kata per kata dari kata yang sering muncul di visualisasi *WordCloud*, kemudian untuk validasi sentimen akan dilakukan identifikasi gabungan kata untuk dilakukan validasi sehingga memperoleh sentimen yang tepat.

4.3.1 *Cluster 1*

a. Visualisasi *WordCloud*

Pada visualisai *WordCloud* ini, ditampilkan kata-kata atau topik yang paling sering dibahas. Gambar 4. 16 **Error! Reference source not found.** merupakan tampilan informasi yang paling sering muncul dibahas oleh konsumen pada komentar konten mengenai Kebijakan Sertifikasi Halal pada penelitian ini.



Gambar 4. 16 Visualisasi *WordCloud Cluster 1*

Dari visualisasi *cluster 1* di atas, dapat dilihat bahwa hasil kata paling sering dibicarakan atau yang dominan seperti kata ‘tidak’, ‘halal’, ‘agama’, ‘Indonesia’, dan lainnya memiliki frekuensi tertentu dengan kata huruf paling besar maka paling banyak dibicarakan, begitu pula sebaliknya.

b. Frekuensi dan Pembobotan Kata

Pada visualisasi menggunakan *WordCloud* diperoleh kata dengan frekuensi tertinggi sebanyak 19 kata dengan bobot dan total nya sebagai berikut:

Tabel 4. 5 Frekuensi dan Pembobotan Kata

Kata Dominan	Frekuensi Kata	Bobot Kata	Total Bobot Kata
tidak	80	-1	-80
bapak	76	0	0
halal	75	+1	+75
saja	56	0	0
agama	39	0	0
lagi	38	0	0
orang	36	0	0

Kata Dominan	Frekuensi Kata	Bobot Kata	Total Bobot Kata
kalau	36	0	0
bisa	35	+1	+35
Indonesia	34	0	0
negara	32	0	0
aku	32	0	0
sudah	32	0	0
cuan	32	+1	+32
nya	32	0	0
ya	31	0	0
harus	31	0	0
apa	28	0	0
seperti	28	0	0
TOTAL BOBOT			62

Dari hasil kata dominan di atas pada *cluster* 1 diperoleh nilai total bobot kata yaitu +62 yang menandakan bahwa *cluster* ini relatif condong ke sentimen positif. Dengan kata positif diantaranya ‘halal’, ‘bisa’, ‘cuan’. Dan kata negatif diantaranya ‘tidak’, untuk kata dengan sentimen netral yaitu sisanya, maka ada kata ‘bapak’, ‘halal’, ‘agama’, ‘lagi’, ‘orang’, ‘kalau’, ‘Indonesia’, ‘negara’, ‘aku’, ‘sudah’, ‘nya’, ‘ya’, ‘harus’, ‘apa’, ‘seperti’. Dari kata yang didapat dari *cluster* 1 ini, maka seharusnya *cluster* ini akan dianggap positif, namun dalam suatu komentar utuh dalam 1 kalimat maka belum tentu bahwa komentar tersebut dapat dikatakan positif hanya dengan 1 kata, untuk itu perlu validasi lebih lanjut dengan frekuensi 2 kata.

c. Frekuensi dan Pembobotan Kata Gabungan

Dari hasil kata gabungan dominan di atas pada *cluster* 1 diperoleh nilai total bobot kata gabungan yaitu -28 yang menandakan bahwa *cluster* ini relatif condong ke sentimen negatif. Dengan kata gabungan positif diantaranya ‘bhineka tunggal’, ‘tunggal ika’, ‘halal bapak’, ‘negara kita’, ‘asal gratis’. Dan kata gabungan negatif diantaranya ‘bukan negara’, ‘cuan cuan’, ‘tidak ada’, ‘makan haram’, ‘cuan lagi’, ‘cari cuan’, ‘agak laen’, ‘tidak halal’. Untuk kata dengan sentimen netral yaitu sisanya, maka ada kata ‘seperti gini’, ‘negara islam’, ‘undang undang’, dan lainnya. Dari kata gabungan yang didapat dari *cluster* 1

cenderung dianggap negatif, dengan representasi bahwa topik kritik terhadap kebijaksanaan ertifikasi halal yang dianggap memiliki kaitan dengan kepentingan ekonomi yang ditunjukkan oleh kata cuan, haram, agama, islam, bhineka tunggal ika, dan lainnya yang membentuk representasi berupa sindiran atau sarkasme terhadap konteks agama juga.

Tabel 4. 6 Frekuensi dan Pembobotan Kata Gabungan

Kata Gabungan Dominan		Frekuensi Kata	Bobot Kata	Total Bobot Kata
bukan	negara	13	-1	-13
cuan	cuan	11	-1	-11
seperti	gini	9	0	0
negara	islam	7	0	0
tidak	ada	7	-1	-7
makan	haram	6	-1	-6
undang	undang	5	0	0
bhineka	tunggal	5	+1	5
agama	islam	5	0	0
cuan	lagi	5	-1	-5
tunggal	ika	5	+1	5
negara	agama	4	0	0
cari	cuan	4	-1	-4
halal	bapak	4	+1	4
negara	kita	4	+1	4
asal	gratis	4	+1	4
agak	laen	4	-1	-4
tidak	halal	4	-1	-4
aku	sudah	4		0
TOTAL BOBOT				-28

Identifikasi kata gabungan tersebut menentukan validasi terhadap kata per kata yang muncul pada *cluster* 1, karena jika ditentukan hanya setiap kata saja maka sentimen belum selalu tepat, untuk itu gabungan kata ini menjadi validasi terhadap setiap komentar pada

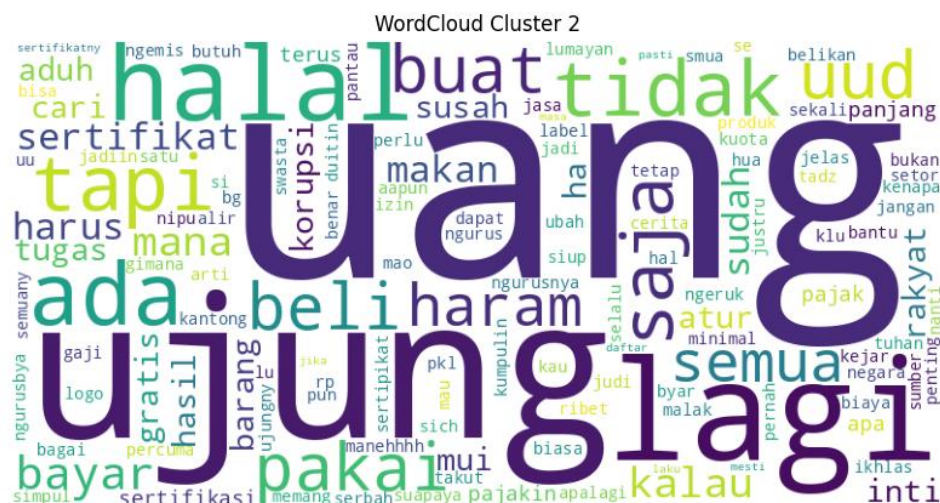
cluster 1 ini. Berikut merupakan contoh kalimat utuh dari sebagian *cluster* 1 berdasarkan pembobotan kata:

1. Berarti anda TIDAK ada Toleransi antar umat Beragama. Indonesia ini sangat beragam (beda suku, adat, Ras, Bahasa, AGAMA, dll) Indonesia BUKAN negara AGAMA TERTENTU. Indonesia itu Bhineka tunggal Ika.
2. cuan cuan cuan...
3. ampun mentri seperti gini kok bisa jadi
4. ini bukan negara Islam Pak, hello mungkin bpk lagi mimpikah?? saya orang Manado Pak saya rakus makan babi, tikus hutan, Anjing, kucing hutan bagaimana ya pakðŸ~□ðŸ~
5. Berarti anda TIDAK ada Toleransi antar umat Beragama. Indonesia ini sangat beragam (beda suku, adat, Ras, Bahasa, AGAMA, dll) Indonesia BUKAN negara AGAMA TERTENTU. Indonesia itu Bhineka tunggal Ika.

4.3.2 Cluster 2

a. Visualisasi WordCloud

Pada visualisai *WordCloud cluster* 2 ini, ditampilkan kata-kata atau topik yang paling sering dibahas. Gambar 4. 17 merupakan tampilan informasi yang paling sering muncul dibahas oleh konsumen pada komentar konten mengenai Kebijakan Sertifikasi Halal pada penelitian ini.



Gambar 4. 17 Visualisasi WordCloud *Cluster* 2

Dari visualisasi *cluster* 2 di atas, dapat dilihat bahwa hasil kata paling sering dibicarakan atau yang dominan seperti kata ‘uang’, ‘ujung’, ‘halal’, ‘bayar’, dan

lainnya memiliki frekuensi tertentu dengan kata huruf paling besar maka paling banyak dibicarakan, begitu pula sebaliknya.

b. Frekuensi dan Pembobotan Kata

Pada visualisasi menggunakan *WordCloud* diperoleh kata dengan frekuensi tertinggi sebanyak 20 kata dengan bobot dan total nya sebagai berikut:

Tabel 4. 7 Frekuensi dan Pembobotan Kata

Kata Dominan	Frekuensi Kata	Bobot Kata	Total Bobot Kata
uang	69	0	0
ujung	25	0	0
lagi	21	0	0
halal	13	+1	13
nya	11	0	0
ada	9	0	0
tapi	7	0	0
tidak	7	-1	-7
saja	6	0	0
pakai	6	0	0
buat	6	0	0
uud	6	0	0
beli	6	0	0
haram	5	-1	-5
semua	5	0	0
bayar	5	0	0
mana	4	0	0
kalau	4	0	0
sertifikat	35	+1	35
harus	3	0	0
TOTAL BOBOT			36

Dari hasil kata dominan di atas pada *cluster* 2 diperoleh nilai total bobot kata yaitu +36 yang menandakan bahwa *cluster* ini relatif condong ke sentimen positif.

Dengan kata positif diantaranya ‘halal’, ‘sertifikat’. Dan kata negatif diantaranya ‘tidak’, ‘haram’. Untuk kata dengan sentimen netral yaitu sisanya, maka ada kata ‘uang’, ‘ujung’, ‘lagi’, ‘lagi’, dan kata lainnya dengan bobot 0. Dari kata yang didapat dari *cluster* 2 ini, maka seharusnya *cluster* ini akan dianggap positif, namun dalam suatu komentar utuh dalam 1 kalimat maka belum tentu bahwa komentar tersebut dapat dikatakan positif hanya dengan 1 kata, untuk itu perlu validasi lebih lanjut dengan frekuensi 2 kata.

c. Frekuensi dan Pembobotan Kata Gabungan

Dari hasil kata gabungan dominan di atas pada *cluster* 2 diperoleh nilai total bobot kata gabungan yaitu -28 yang menandakan bahwa *cluster* ini relatif condong ke sentimen negatif. Dengan kata gabungan positif diantaranya ‘uang halal’. Dan kata gabungan negatif diantaranya ‘uang lagi’, ‘uang uang’, ‘ujung uang’, ‘pakai uang’, ‘uang haram’, ‘uang haram’, ‘halal tapi’, ‘lagi uang’, ‘makan uang’, ‘cari uang’, ‘beli harus’. Untuk kata dengan sentimen netral yaitu sisanya, maka ada kata ‘ujung ujung’, ‘uud ujung’, ‘lagi saja’, dan lainnya. Dari kata gabungan yang didapat dari *cluster* 2 cenderung dianggap negatif, dengan representasi bahwa topik kritik terhadap kebijakans sertifikasi halal ini dianggap bahwa konsumen memiliki ketidakpercayaan terhadap kehalalan, terutama karena adanya persepsi bahwa aspek finansial seperti mahal, pungutan, dan syarat tertentu yang bertentangan dengan kehalalan.

Tabel 4. 8 Frekuensi dan Pembobotan Kata Gabungan

Kata Gabungan Dominan		Frekuensi Kata	Bobot Kata	Total Bobot Kata
uang	lagi	14	-1	-14
ujung	ujung	9	0	0
uang	uang	8	-1	-8
ujung	uang	6	-1	-6
uud	ujung	5	0	0
lagi	saja	4	0	0
pakai	uang	4	-1	-4
uang	haram	4	-1	-4
uang	nya	3	0	0

Kata Gabungan Dominan		Frekuensi Kata	Bobot Kata	Total Bobot Kata
halal	tapi	3	-1	-3
lagi	uang	3	-1	-3
ujung	nya	3	0	0
makan	uang	2	-1	-2
cari	uang	2	-1	-2
buat	beli	2	0	0
lagi	lagi	2	0	0
beli	harus	2	-1	-2
uang	halal	2	+1	+1
nya	uang	2	0	0
ada	uang	2	0	0
TOTAL BOBOT				-47

Identifikasi kata gabungan tersebut menentukan validasi terhadap kata per kata yang muncul pada *cluster 2*, karena jika ditentukan hanya setiap kata saja maka sentimen belum selalu tepat, untuk itu gabungan kata ini menjadi validasi terhadap setiap komentar pada *cluster 2* ini. Berikut merupakan contoh kalimat utuh dari sebagian *cluster 2* berdasarkan pembobotan kata:

1. aduh **uang lagi uang lagi**
2. gratis tapi susah ribet ngurusnya **ujung ujung** bantu tugas tapi ada uang jasa seikhlas nya tapi minimal Rp...
3. **uang uang**
4. uud **ujung uang**
5. santri² pesantren gda label **halal tapi** banyak jd korban boss..mulut aje rame enteeeee... yahanu...

4.3.3 Cluster 3

a. Visialisasi *WordCloud*

Pada visualisai *WordCloud cluster 3* ini, ditampilkan kata-kata atau topik yang paling sering dibahas. Gambar 4. 18 merupakan tampilan informasi yang paling

Kata Dominan	Frekuensi Kata	Bobot Kata	Total Bobot Kata
Tapi	25	0	0
Kalau	25	0	0
Logo	22	+1	22
Kandung	21	0	0
Saja	21	0	0
Ya	21	0	0
Non	20	-1	-20
Makan	17	+1	17
label	17	+1	17
TOTAL BOBOT			296

Dari hasil kata dominan di atas pada *cluster* 2 diperoleh nilai total bobot kata yaitu +296 yang menandakan bahwa *cluster* ini relatif condong ke sentimen positif. Dengan kata positif diantaranya ‘halal’, ‘sertifikat’. Dan kata negatif diantaranya ‘babi’, ‘tidak’. Untuk kata dengan sentimen netral yaitu sisanya, maka ada kata ‘ada’, ‘harus’ dan kata lainnya dengan bobot 0 pada Tabel 4. 9. Dari kata yang didapat dari *cluster* 3 ini, maka seharusnya *cluster* ini akan dianggap positif, namun dalam suatu komentar utuh dalam 1 kalimat maka belum tentu bahwa komentar tersebut dapat dikatakan positif hanya dengan 1 kata, untuk itu perlu validasi lebih lanjut dengan frekuensi 2 kata.

c. Frekuensi dan Pembobotan Kata Gabungan

Dari hasil kata gabungan dominan di atas pada *cluster* 3 diperoleh nilai total bobot kata gabungan yaitu 108 yang menandakan bahwa *cluster* ini relatif condong ke sentimen positif. Dari kata gabungan yang didapat dari *cluster* 3 cenderung dianggap positif, dengan representasi bahwa konsumen memiliki kesadaran tinggi akan pentingnya sertifikasi halal, namun masih disertai keraguan dan kekhawatiran terhadap kehalalan produk, terutama terkait kandungan non-halal seperti babi.

Tabel 4. 10 Frekuensi dan Pembobotan Kata Gabungan

Kata Gabungan Dominan		Frekuensi Kata	Bobot Kata	Total Bobot Kata
sertifikat	halal	74	+1	74
logo	halal	19	+1	19
non	halal	17	-1	-17
harus	ada	16	0	0
ada	sertifikat	15	+1	15
kandung	babi	15	-1	-15
label	halal	13	+1	13
halal	nya	12	+1	12
halal	tidak	12	-1	-12
halal	tapi	12	-1	-12
ada	logo	12	0	0
label	halal	22	+1	22
halal	ya	12	+1	12
halal	tidak	12	-1	-12
halal	tapi	12	-1	-12
ada	logo	12	0	0
halal	ya	9	+1	+9
harus	sertifikat	9	+1	+9
wajib	sertifikat	8	+1	+8
harus	halal	8	+1	+8
babi	harus	7	-1	-7
daging	babi	6	-1	-6
ada	label	6	0	0
tapi	ada	6	0	0
TOTAL BOBOT				108

Identifikasi kata gabungan tersebut menentukan validasi terhadap kata per kata yang muncul pada *cluster* 3, karena jika ditentukan hanya setiap kata saja maka sentimen belum selalu tepat, untuk itu gabungan kata ini menjadi validasi terhadap

Tabel 4. 11 Frekuensi dan Pembobotan Kata

Kata Dominan	Frekuensi Kata	Bobot Kata	Total Bobot Kata
dagang	47	0	0
tidak	23	-1	-23
kaki	16	0	0
kalau	16	0	0
sertifikat	13	+1	13
halal	13	+1	13
lima	13	0	0
babi	10	-1	-10
gimana	9	0	0
aku	8	0	0
kakak	8	0	0
jual	8	0	0
tapi	7	0	0
punya	7	0	0
harus	7	0	0
nya	6	0	0
non	6	-1	-6
ada	6	0	0
gratis	6	+1	6
saja	6	0	0
TOTAL BOBOT			-7

Dari hasil kata dominan di atas pada *cluster* 4 diperoleh nilai total bobot kata yaitu -7 yang menandakan bahwa *cluster* ini relatif condong ke sentimen negatif tetapi masih dalam sekumpulan kata netral. Dengan kata positif diantaranya ‘halal’, ‘sertifikat’. Dan kata negatif diantaranya ‘non’, ‘babi’. Untuk kata dengan sentimen netral yaitu sisanya, maka ada kata ‘gimana’, ‘aku’, dan kata lainnya dengan bobot 0 pada Tabel 4. 11. Dari kata yang didapat dari *cluster* 5 ini, maka seharusnya *cluster* ini akan dianggap negatif tetapi memiliki kecenderungan

terhadap kata netral juga karena sebagian besar juga bukan berisi opini langsung melainkan sebuah pertanyaan, namun dalam suatu komentar utuh dalam 1 kalimat maka belum tentu bahwa komentar tersebut dapat dikatakan sentimen tersebut hanya dengan 1 kata, untuk itu perlu validasi lebih lanjut dengan frekuensi 2 kata.

c. Frekuensi dan Pembobotan Kata Gabungan

Dari hasil kata gabungan dominan di atas pada *cluster* 4 diperoleh nilai total bobot kata gabungan yaitu +9 yang menandakan bahwa *cluster* ini relatif condong ke sentimen positif, namun masih dalam sekumpulan kata netral. Dari kata gabungan yang didapat dari *cluster* 4 cenderung dianggap positif cenderung netral ini, dengan representasi bahwa konsumen atau UMKM memiliki pertanyaan seputar kewajiban sertifikasi halal, terutama pada pedagang kecil (UMKM) yang masih membutuhkan penjelasan dan masih diskusi satu sama lain antar komentar.

Tabel 4. 12 Frekuensi dan Pembobotan Kata Gabungan

Kata Gabungan Dominan		Frekuensi Kata	Bobot Kata	Total Bobot Kata
kaki	lima	12	0	0
dagang	kaki	10	0	0
sertifikat	halal	7	+1	7
kalau	dagang	6	0	0
punya	sertifikat	5	+1	5
harus	punya	4	0	0
non	halal	3	-1	-3
dagang	makan	2	0	0
daging	babi	2	0	0
kalau	tidak	2	-1	-2
babi	bisa	2	-1	-2
sate	babi	2	0	0
babi	tidak	2	0	0
orang	dagang	2	+1	2
halal	tidak	2	0	0
nya	tidak	2	0	0
sertifikat	tidak	2	0	0

Kata Gabungan Dominan		Frekuensi Kata	Bobot Kata	Total Bobot Kata
setuju	tapi	2	0	0
dapat	sertifikat	2	+1	2
TOTAL BOBOT				+9

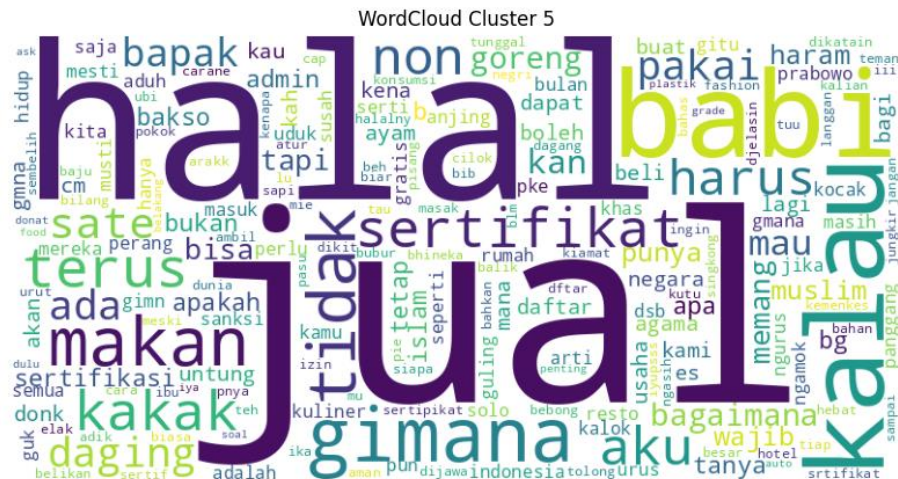
Identifikasi kata gabungan tersebut menentukan validasi terhadap kata per kata yang muncul pada *cluster* 4, karena jika ditentukan hanya setiap kata saja maka sentimen belum selalu tepat, untuk itu gabungan kata ini menjadi validasi terhadap setiap komentar pada *cluster* 4 ini. Berikut merupakan contoh kalimat utuh dari sebagian *cluster* 4 berdasarkan pembobotan kata:

1. di sekolah gw juga pedagang yang mau masuk juga harus punya **sertifikat halal**, tapi pas ngurus juga dibantu sama pihak sekolah
2. **kalau dagang** baju pakai sertifikat gak?
3. kalo yg dagang makanan **non-halal** gmn?
4. dagang sate **babi bisa** dapet sertifikat halal ga bang? ðŸ—¿
5. bagus, jadi ga sembarangan **orang dagang** jadi tau mana orang dagang niat dan tidak yg tidak pasti akan bilng ribet ah

4.3.5 Cluster 5

a. Visialisasi *WordCloud*

Pada visualisai *WordCloud cluster* 5 ini, ditampilkan kata-kata atau topik yang paling sering dibahas. Gambar 4. 20 merupakan tampilan informasi yang paling sering muncul dibahas oleh konsumen pada komentar konten mengenai Kebijakan Sertifikasi Halal pada penelitian ini.



Gambar 4. 20 Visualisasi WordCloud *Cluster 5*

Dari visualisasi *cluster 5* di atas, dapat dilihat bahwa hasil kata sering dibicarakan atau yang dominan seperti kata ‘halal, ‘jual, ‘makan, ‘sertifikat, dan lainnya memiliki frekuensi tertentu dengan kata huruf paling besar maka paling banyak dibicarakan, begitu pula sebaliknya.

b. Frekuensi dan Pembobotan Kata

Pada visualisasi menggunakan *WordCloud* diperoleh kata dengan frekuensi tertinggi sebanyak 20 kata dengan bobot dan total nya sebagai berikut:

Tabel 4. 13 Frekuensi dan Pembobotan Kata

Kata Dominan	Frekuensi Kata	Bobot Kata	Total Bobot Kata
jual	86	0	0
halal	60	+1	60
babi	56	-1	-56
kalau	44	-1	0
gimana	43	0	0
makan	34	0	0
tidak	33	-1	-33
terus	23	0	0
sertifikat	22	+1	22
kakak	22	0	0
aku	20	0	0
non	19	-1	-19

Kata Dominan	Frekuensi Kata	Bobot Kata	Total Bobot Kata
harus	17	0	0
sate	14	0	0
pakai	12	0	0
daging	12	0	0
ada	12	0	0
bapak	11	0	0
mau	9	0	0
ya	9	0	0
TOTAL BOBOT			-26

Dari hasil kata dominan di atas pada *cluster* 5 diperoleh nilai total bobot kata yaitu -26 yang menandakan bahwa *cluster* ini relatif condong ke sentimen negatif tetapi masih dalam sekumpulan kata netral. Dengan kata positif diantaranya ‘halal’, ‘sertifikat’. Dan kata negatif diantaranya ‘kalau’, ‘babi’, ‘tidak’, dan lainnya. Untuk kata dengan sentimen netral yaitu sisanya, maka ada kata ‘gimana’, ‘jual’, ‘makan’, dan kata lainnya dengan bobot 0 pada tabel Tabel 4. 13. Dari kata yang didapat dari *cluster* 5 ini, maka seharusnya *cluster* ini akan dianggap negatif tetapi memiliki kecenderungan terhadap kata netral juga karena sebagian besar juga bukan berisi opini langsung melainkan sebuah pertanyaan, namun dalam suatu komentar utuh dalam 1 kalimat maka belum tentu bahwa komentar tersebut dapat dikatakan sentimen tersebut hanya dengan 1 kata, untuk itu perlu validasi lebih lanjut dengan frekuensi 2 kata.

c. Frekuensi dan Pembobotan Kata Gabungan

Dari hasil kata gabungan dominan di atas pada *cluster* 4 diperoleh nilai total bobot kata gabungan yaitu -41 yang menandakan bahwa *cluster* ini relatif condong ke sentimen negatif. Dari kata gabungan yang didapat dari *cluster* 4 cenderung dianggap negatif, dengan representasi bahwa persepsi konsumen terkait isu kehalalan produk ini ditunjukkan oleh dominasi kata-kata ‘tidak halal’, ‘jual non’, ‘babi’, dan kombinasi kata lainnya yang mengarah pada pelanggaran prinsip halal. Walaupun terdapat kata positif seperti ‘sertifikat halal’. Namun, *cluster* 5 ini lebih mengarah pada pembicaraan terkait harapan atau tuntutan terhadap perbaikan,

bukan sentimen utama. Jadi, *cluster* ini merepresentasikan kekhawatiran, kritik, serta ketidakpercayaan konsumen terhadap produk yang beredar.

Tabel 4. 14 Frekuensi dan Pembobotan Kata Gabungan

Kata Gabungan Dominan		Frekuensi Kata	Bobot Kata	Total Bobot Kata
kalau	jual	15	0	0
babi	gimana	14	0	0
non	halal	13	-1	-13
jual	sate	11	0	0
jual	makan	11	0	0
sertifikat	halal	10	+1	10
daging	babi	9	-1	-9
makan	babi	9	0	0
Aku	jual	9	0	0
sate	babi	9	-1	-9
halal	gimana	9	0	0
jual	babi	8	-1	-8
makan	non	7	-1	-7
jual	daging	7	0	0
tidak	halal	6	-1	-6
pakai	sertifikat	6	+1	6
gimana	kakak	6	0	0
terus	jual	6	0	0
jual	non	5	-1	-5
non	Muslim	4	0	0
TOTAL BOBOT				-41

Identifikasi kata gabungan tersebut menentukan validasi terhadap kata per kata yang muncul pada *cluster* 5, karena jika ditentukan hanya setiap kata saja maka sentimen belum selalu tepat, untuk itu gabungan kata ini menjadi validasi terhadap setiap komentar pada *cluster* 5 ini. Berikut merupakan contoh kalimat utuh dari sebagian *cluster* 5 berdasarkan pembobotan kata:

1. Bang kalau jualannya emang **tidak halal** juga bakal dikenai sanksi?

2. Yg jual **non halal** mesti ada sertifikat haramðŸŒŸðŸŒŸ
3. ngak masuk di akal sehat bpk, kalau semuanya di halalkan, trus bagaiman kita non Muslim, **babi** di halalkan
4. kalau buat jual jajan seperti babi guling dsb memang nikmat non **gimana kakak** masa mereka tidak boleh makan makan mereka
5. lha kalo pedaganh kaki lima yg **jual non** halal bgmn?

4.4 Hasil Ringkasan Ciri Khas Sentimen tiap-tiap Cluster

Setelah dilakukan identifikasi ciri khas setiap *cluster* kemudian dilanjutkan dengan identifikasi faktor kemungkinan penyebab dari masing-masing *cluster*, dari perwakilan komentar untuk menentukan sentimen yang didapatkan tiap *cluster*.

Tabel 4. 15 Ciri Khas Tiap Cluster

<i>Cluster</i>	Fakotr Kemungkinan Penyebab	Sentimen	Label Sentimen
Cluster 1	Persepsi konsumen terhadap sertifikasi halal ini cenderung menganggap kebijakan ini sebagai sarana untuk memperoleh keuntungan ekonomi, dengan contoh kata sering dibahas yaitu kata ‘ cuan cuan ’ dan ‘ cuan lagi ’	Negatif 	<i>Profit-Oriented perception</i> (persepsi keuntungan ekonomi)
Cluster 2	Topik kritik terhadap kebijakan sertifikasi halal ini dianggap bahwa konsumen memiliki ketidakpercayaan atau penolakan terhadap kebijakan ini, terutama karena adanya persepsi	Negatif 	<i>Cost Resistance</i> (resistensi biaya)

<i>Cluster</i>	Fakotr Kemungkinan Penyebab	Sentimen	Label Sentimen
	bahwa aspek finansial seperti mahal, pungutan, dan syarat tertentu yang bertentangan dengan kehalalan, seperti contoh kata sering dibahas yaitu ‘ uang haram ’, ‘ halal tapi ’, ‘ lagi uang ’, ‘ makan uang ’		
<i>Cluster 3</i>	Konsumen memiliki kesadaran tinggi akan pentingnya sertifikasi halal, karena masih disertai keraguan dan kekhawatiran terhadap kehalalan produk, terutama terkait kandungan non-halal seperti babi. Dalam <i>cluster</i> ini kata sering dibahas ditunjukkan oleh kata ‘ ada sertifikat ’, ‘ label halal ’, ‘ wajib sertifikat ’	Positif 	<i>Halal Awarness</i> (kesadaran halal)
<i>Cluster 4</i>	Pada <i>cluster</i> ini, konsumen atau UMKM memiliki banyaknya pertanyaan seputar kewajiban sertifikasi	Netral 	<i>Regulation Confusion</i> (kebingungan regulasi/kebijakan)

<i>Cluster</i>	Fakotr Kemungkinan Penyebab	Sentimen	Label Sentimen
	halal, terutama pada pedagang kecil (UMKM) yang masih membutuhkan penjelasan dan masih diskusi satu sama lain. Contoh kata sering dibahas yaitu kata ‘Kaki lima’, ‘dagang kaki’, ‘harus punya’, ‘sertifikat tidak’		
<i>Cluster 5</i>	Persepsi konsumen terkait isu kehalalan produk ditunjukkan oleh dominasi kata-kata ‘tidak halal’, ‘jual non’, ‘babi’, dan kombinasi kata lainnya yang mengarah pada pelanggaran prinsip halal. Namun, <i>cluster 5</i> ini lebih mengarah pada pembicaraan terkait harapan atau tuntutan terhadap perbaikan. Jadi, <i>cluster</i> ini merepresentasikan kekhawatiran, kritik, serta ketidakpercayaan konsumen terhadap produk yang beredar,	Negatif 	<i>Halal Concern</i> (kekhawatiran kehalalan produk)

<i>Cluster</i>	Faktor Kemungkinan Penyebab	Sentimen	Label Sentimen
	karena mereka khawatir akan produk yang seharusnya haram tetapi karena kebijakan ini, mereka melabelinya dengan label halal, terutama untuk konsumen Muslim.		

Pada Tabel 4. 1 menunjukkan proporsi sentimen dalam *cluster* dalam penelitian ini.

Tabel 4. 16 Proporsi Sentimen

Sentimen	Jumlah	Persentase
Negatif	693	72,11%
Positif	161	16,75%
Netral	107	11,13%

Pada proporsi sentimen ini, terdapat jumlah 693 data berlabel negatif dengan persentase 72,11%. Lalu terdapat jumlah 161 data berlabel positif dengan persentase 16,75%, dan selanjutnya terdapat jumlah 107 data berlabel netral dengan persentase 11,13%. Dari jumlah tersebut, *cluster* dalam penelitian ini memiliki sentimen terbesar dengan label negatif sebesar 72,11% data, maka dari itu *cluster* dengan sentimen ini memiliki analisis dan perbaikan paling diperlukan.

BAB V

PEMBAHASAN

5.1 Profiling *Cluster*

Merujuk pada hasil pengolahan data pada tahap sebelumnya, jumlah *cluster* yang optimal yaitu lima *cluster* berdasarkan metode elbow yang sudah diterapkan. Lalu dari perolehan *cluster* tersebut kemudian diolah menggunakan metode *lexicon* di *google colab* dengan kode tertentu menggunakan bahasa pemrograman *python*, dengan demikian hasil tersebut digunakan untuk mendapatkan analisis ciri khas dari setiap *cluster* nya, berikut merupakan uraian singkat label perolehan ciri khas tiap-tiap *cluster* dari penelitian ini.

5.1.1 *Cluster 1*

Merujuk pada hasil pengolahan data menggunakan metode algoritma K-Means untuk proses *clustering*, dan menggunakan *lexicon* untuk proses mendapatkan sentimen, maka *cluster 1* cenderung menunjukkan opini negatif. Berdasarkan hasil visualisasi *wordcloud* dan pembobotan TF-IDF, kata-kata yang sering dibicarakan pada *cluster* ini cenderung negatif. Makna dari dominasi kata ‘cuan’ dan ‘uang’ menunjukkan bahwa masyarakat memiliki persepsi bahwa kebijakan sertifikasi halal dimanfaatkan sebagai sarana ekonomi atau komersial oleh pihak tertentu. Maka, *Cluster 1* memiliki representasi kelompok masyarakat dengan *profit-oriented perception* atau persepsi yang menganggap kebijakan sertifikasi halal tidak hanya memiliki kaitan dengan aspek agama dan perlindungan konsumen, tetapi juga memiliki unsur kepentingan finansial dalam penerapannya.

5.1.2 *Cluster 2*

Merujuk pada hasil pengolahan data menggunakan metode algoritma K-Means untuk proses *clustering*, dan menggunakan *lexicon* untuk proses mendapatkan sentimen, maka *cluster 2* cenderung menunjukkan opini negatif. Berdasarkan hasil visualisasi *wordcloud* dan pembobotan TF-IDF. Makna dari dominasi kata yang sering dibicarakan ini, memiliki arti bahwa kata ‘uang’ dan ‘lagi’, mengindikasikan bahwa masyarakat menganggap bahwa pemerintah selalu membuat kebijakan dengan pembebanan biaya dalam pembuatan sertifikasi halal atau *cost resistance* (resistensi biaya).

5.1.3 *Cluster 3*

Merujuk pada hasil pengolahan data menggunakan metode algoritma K-Means untuk proses *clustering*, dan menggunakan *lexicon* untuk proses mendapatkan sentimen, maka

cluster 3 cenderung menunjukkan opini positif. Berdasarkan hasil visualisasi *wordcloud* dan pembobotan TF-IDF, kata-kata yang sering dibicarakan pada *cluster* ini cenderung positif. Makna dari dominasi kata yang sering dibicarakan ini, memiliki arti bahwa kata ‘harus’ dan ‘halal’, mengindikasikan bahwa masyarakat menganggap kebijakan sertifikasi halal ini sebagai penerapan yang penting atau *halal awarness* (kesadaran halal). Di sisi lain karena mereka memiliki kekhawatiran terhadap produk yang tidak halal, terutama bagi masyarakat Muslim terkait produk dengan kandungan non-halal yaitu seperti daging babi.

5.1.4 *Cluster 4*

Merujuk pada hasil pengolahan data menggunakan metode algoritma K-Means untuk proses *clustering*, dan menggunakan *lexicon* untuk proses mendapatkan sentimen, maka *cluster* 4 cenderung menunjukkan opini netral. Berdasarkan hasil visualisasi *wordcloud* dan pembobotan TF-IDF, makna dari dominasi kata yang sering dibicarakan ini, memiliki arti bahwa kata ‘dagang dan ‘kaki lima’, mengindikasikan bahwa masyarakat pada *cluster* ini konsumen atau UMKM memiliki banyaknya pertanyaan seputar kewajiban sertifikasi halal, terutama pada pedagang kecil (UMKM) yang masih membutuhkan penjelasan dan masih diskusi satu sama lain. atau *Regulation Confusion* (kebingungan regulasi/kebijakan).

5.1.5 *Cluster 5*

Cluster 5 merupakan kelompok komentar dengan dominasi sentimen negatif terhadap kebijakan sertifikasi halal di Indonesia. Merujuk pada hasil pengolahan data menggunakan metode algoritma K-Means untuk proses *clustering*, dan menggunakan *lexicon* untuk proses mendapatkan sentimen, maka *cluster* ini cenderung menunjukkan opini negatif. Berdasarkan hasil visualisasi *wordcloud* dan pembobotan TF-IDF, kata-kata yang sering dibicarakan pada *cluster* ini cenderung negatif, makna dari dominasi kata yang sering dibicarakan ini, memiliki arti bahwa kata ‘jual babi’ dan ‘makan non’, mengindikasikan bahwa masyarakat pada *cluster* ini khawatir akan kehalalan produk non-halal apabila kebijakan ini diterapkan tetapi tidak sesuai dengan semestinya, seperti produk yang seharusnya non-halal, namun akan berlabelkan halal, yang mengarah pada pelanggaran prinsip halal.

5.2 Desain Usulan Strategi Percepatan Sertifikasi Halal

Merujuk pada hasil profiling *cluster* 1 hingga *cluster* 5 yang hanya menunjukkan *cluster* yang masih perlu penanganan seperti *cluster* negatif dan netral. Pada bagian sebelumnya, yang memiliki ciri khas tiap *cluster*nya, maka langkah selanjutnya yaitu membuat desain usulan strategi percepatan sertifikasi halal untuk setiap *cluster* yang sudah dibentuk. Tabel 5. 1 merupakan strategi percepatan sertifikasi halal berdasarkan *cluster* opini yang terbentuk.

Tabel 5. 1 Strategi Percepatan Sertifikasi Halal

No.	Strategi	Referensi	Cluster 1	Cluster 2	Cluster 4	Cluster 5	Rank
1	Sosialisasi transparansi biaya menggunakan alat berbasis teknologi informasi (kalkulator BPJPH)	(Kamela, 2026), (Yan et al., 2020)	√	√			2
2	Kampanye testimoni UMKM yang berhasil meningkatkan penjualan setelah sertifikasi halal	(N. Azizah et al., 2025), (Patel, 2024)		√	√		3
3	Kolaborasi pemerintah dengan <i>influencer</i> atau pelaku usaha halal dalam meningkatkan	(Razali & Samori, 2025), (Ningsih et al., 2025)	√	√	√	√	1

No.	Strategi	Referensi	Cluster 1	Cluster 2	Cluster 4	Cluster 5	Rank
	literasi digital sertifikasi halal melalui konten edukatif TikTok.						
4	Program Kampanye “Kenali Bahan non-Halal” untuk pemahaman masyarakat.	(Putro et al., 2025), (Q. Azizah & Dzikrulloh, 2024)			√	√	4

Berikut merupakan penjelasan dari implementasi desain strategi yang diusulkan berdasarkan hasil *Cluster 1*, *Cluster 2*, *Cluster 4*, dan *Cluster 5*:



Gambar 5. 1 Implementasi Desain Komunikasi terhadap Publik

Pada Gambar 5. 1 tersebut menjelaskan proses atau alur implementasi dari penyampaian informasi kepada publik atau masyarakat terkait desain konkrit dari strategi percepatan sertifikasi halal berdasarkan profiling cluster yang telah diketahui. Desain alur penyampaian informasi ini disusun dari hasil analisis cluster kemudian ditemukannya *information gap* yang terjadi antara persepsi masyarakat terhadap fakta yang dimiliki oleh BPJPH. Selanjutnya BPJPH dapat memberikan materi edukasi atau program berdasarkan strategi usulan yang dibuat berdasarkan penelitian ini. Terdapat empat strategi usulan yang memiliki contoh implementasi program strategi. Berikut merupakan penjelasan dari strategi yang diusulkan berdasarkan tabel Tabel 5. 1 di atas:

5.2.1 Strategi 1 Sosialisasi Transparansi Biaya

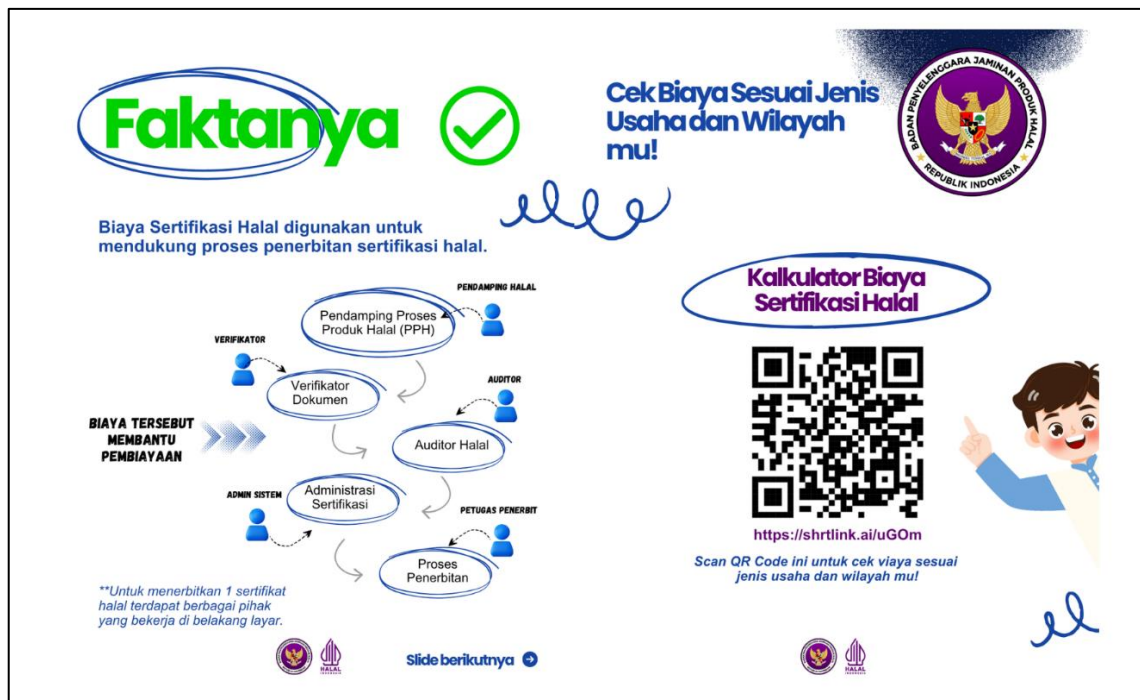
Hasil analisis Cluster 1 dan Cluster 2 menunjukkan bahwa masih terdapat persepsi negatif masyarakat terhadap biaya sertifikasi halal. Pada Cluster 1 yang mencakup 56,82% opini memiliki persepsi bahwa sertifikasi halal ini sebagai sarana untuk memperoleh keuntungan ekonomi, dan *cluster 2* sebanyak 6,04% opini memiliki persepsi terhadap aspek finansial seperti mahal, pungutan, dan syarat tertentu yang bertentangan dengan kehalalan. Persepsi tersebut menunjukkan bahwa sebagian masyarakat masih memahami biaya sertifikasi halal sebagai bentuk pemasukan pemerintah semata.

Padahal, menurut PP Nomor 39 Tahun 2021 tentang Penyelenggaraan Bidang Jaminan Produk Halal Pasal 85 menyatakan bahwa biaya sertifikasi halal dibebankan kepada pelaku usaha dan digunakan dalam penyelenggaraan proses sertifikasi halal. Biaya sertifikasi halal tidak digunakan untuk kepentingan pemerintah secara langsung, melainkan dialokasikan untuk mendukung keseluruhan proses sertifikasi halal, mulai dari pendampingan pelaku usaha, verifikasi dokumen, pemeriksaan dokumen, hingga penerbitan sertifikat halal. Selain itu, biaya tersebut juga digunakan untuk mendukung operasional sumber daya manusia yang terlibat dalam proses sertifikasi, seperti Pendamping Proses Produk Halal (PPPH), auditor halal, verifikator, dan tenaga administrasi. Hal seperti inilah yang menunjukkan adanya kesenjangan informasi antara persepsi masyarakat dan fakta terkait penggunaan biaya sertifikasi halal.

Untuk mengatasi kesenjangan informasi tersebut, diusulkan salah satu contoh sosialisasi digital dengan tema “Rp500.000 digunakan untuk apa?” yang memanfaatkan fitur Kalkulator Biaya Sertifikasi Halal sebagai media edukasi utama. Kampanye dapat disampaikan melalui TikTok berbentuk slide, Instagram Reels, atau poster seperti Gambar 5. 2 dan Gambar 5. 3, yang menjelaskan alokasi biaya sertifikasi halal, pihak-pihak yang terlibat dalam proses sertifikasi, serta cara penggunaan kalkulator BPJPH untuk menghitung estimasi biaya sertifikasi halal.



Gambar 5. 2 Contoh Implementasi Program Strategi 1



Gambar 5. 3 Contoh Implementasi Program Strategi 1 (2)

Gambar tersebut menampilkan informasi mengenai komponen penggunaan biaya sertifikasi halal, pihak yang terlibat dalam proses sertifikasi, serta akses menuju fitur Kalkulator Biaya Sertifikasi Halal BPJPH. Informasi ditampilkan secara sederhana agar mudah dipahami oleh pelaku usaha maupun masyarakat umum.

Strategi ini didukung oleh penelitian yang merujuk pada studi literatur oleh (Kamela, 2026), yang menunjukkan bahwa transparansi dan akuntabilitas biaya pendidikan merupakan faktor utama memengaruhi persepsi mahasiswa. Dan validasi studi yang dilakukan oleh (Yan et al., 2020) menunjukkan transparansi keuangan online membuktikan sebesar 4,3 kali naik dibanding tahun tahun sebelumnya terhadap kepercayaan masyarakat terkait investasi swasta. Melalui strategi ini, masyarakat diharapkan dapat memahami biaya sertifikasi halal digunakan untuk mendukung proses sertifikasi halal dan bukan semata-mata menjadi pemasukan pemerintah. Dengan demikian, persepsi negatif terkait biaya sertifikasi halal dapat berkurang dan kepercayaan masyarakat dapat meningkat.

5.2.2 Strategi 2 Kampanye Testimoni UMKM

Hasil analisis pada Cluster 2 menunjukkan bahwa sebagian masyarakat memandang sertifikasi halal sebagai beban tambahan bagi pelaku usaha karena dianggap menambah biaya dan proses administrasi. Selain itu, ditemukan pula keraguan terhadap manfaat yang diperoleh setelah pelaku usaha memperoleh sertifikasi halal. Persepsi ini menunjukkan

bahwa masyarakat lebih banyak melihat sisi biaya dibandingkan manfaat yang diperoleh dari sertifikasi halal.

Pada kenyataannya, sertifikasi halal tidak hanya berfungsi sebagai pemenuhan regulasi atau kebijakan, tetapi juga untuk meningkatkan daya kepercayaan bagi konsumen, memperluas peluang pasar, serta meningkatkan kepercayaan konsumen, memperluas peluang pasar, serta meningkatkan daya saing produk. Namun, manfaat tersebut belum sepenuhnya diketahui oleh masyarakat karena informasi yang diterima lebih banyak berfokus pada biaya dan proses sertifikasi dibanding dampak positif yang dirasakan oleh pelaku usaha.

Untuk itu, dari kondisi tersebut, diusulkan kampanye digital dengan salah satu tema “UMKM Naik Kelas bersama Sertifikasi Halal” yang menampilkan pengalaman nyata pelaku UMKM sebelum dan sesudah memperoleh sertifikasi halal. Konten dapat berupa video testimoni, wawancara singkat, maupun carousel media sosial yang menjelaskan perubahan setelah memperoleh sertifikasi halal, seperti contoh implementasi strategi melalui program pada Gambar 5. 4 dan Gambar 5. 5.



Gambar 5. 4 Contoh Implementasi Program Strategi 2



Gambar 5. 5 Contoh Implementasi Program Strategi 2(2)

Gambar di atas menampilkan pengalaman nyata pelaku UMKM yang telah memperoleh sertifikat halal, mulai dari kondisi sebelum sertifikasi hingga manfaat yang dirasakan setelah memiliki sertifikasi halal kepada masyarakat.

Strategi ini didukung oleh studi literatur yang merujuk pada penelitian oleh (N. Azizah et al., 2025) yang menunjukkan salah satunya bahwa testimonial memiliki pengaruh positif dan signifikan terhadap keputusan pembelian di UMKM di kota Palopo, dan penelitian oleh (Patel, 2024) yang menunjukkan bahwa studi ini meneliti pola pembelian berbagai industri dipengaruhi oleh ulasan atau testimoni pembeli lain, maka melalui strategi ini, masyarakat diharapkan dapat memahami bahwa sertifikasi halal tidak hanya berkaitan dengan biaya dan kewajiban administratif, tetapi juga memberikan manfaat nyata bagi pengembangan usaha dan peningkatan kepercayaan konsumen.

5.2.3 Strategi 3 Kolaborasi pemerintah dengan *influencer*

Hasil analisis pada *Cluster 1*, *Cluster 2*, *Cluster 4*, dan *Cluster 5* menunjukkan masih banyaknya persepsi kritik, serta informasi yang kurang tepat mengenai sertifikasi halal di media sosial. Temuan tersebut mengindikasikan bahwa informasi resmi yang telah disampaikan oleh BPJPH belum sepenuhnya menjangkau kelompok masyarakat yang aktif menggunakan platform digital, khususnya TikTok.

Fakta menunjukkan bahwa BPJPH telah menyediakan berbagai informasi mengenai sertifikasi halal melalui website bpjph.halal.go.id dan media sosial resmi seperti instagram dengan user @halal.indonesia. Namun, karakteristik pengguna media sosial saat ini cenderung lebih mudah menerima informasi yang disampaikan melalui publik figur, influencer, maupun content creator terkenal yang memiliki kedekatan dengan audiensnya. Hal ini menunjukkan adanya kesenjangan antara penyampaian informasi yang dilakukan oleh BPJPH dengan konsumsi informasi masyarakat di era digital.

Oleh karena itu, diusulkan program edukasi digital dengan tema “1 Menit paham Sertifikasi Halal” yang melibatkan influencer dan content creator yang juga dapat bekerja sama dengan pihak BPJPH seperti auditor BPJPH untuk melakukan *podcast Question and Answer*, atau menjawab pertanyaan netizen (masyarakat) seperti contoh konten Gambar 5. 6.



Gambar 5. 6 Contoh Implementasi Program Strategi 3

Gambar tersebut memperlihatkan contoh konten edukasi yang dikemas secara ringan melalui format video singkat atau cuplikan singkat dari podcast dan pembahasan seperti mitos atau fakta terkait sertifikasi halal. Format video dirancang agar sesuai dengan karakteristik pengguna media sosial yang lebih menyukai informasi yang cepat dan mudah dipahami.

Strategi ini didukung oleh penelitian di bidang komunikasi digital merujuk dari penelitian yang dilakukan oleh (Razali & Samori, 2025) menunjukkan bahwa kolaborasi yang dilakukan dengan influencer Muslim memiliki pengaruh positif terhadap minat beli produk halal dan penelitian oleh (Ningsih et al., 2025) juga memberikan hasil penelitian bahwa influencer media sosial memiliki peran penting dalam *consumer behavior* atau perilaku konsumen. Melalui strategi ini, informasi mengenai sertifikasi halal diharapkan dapat menjangkau lebih banyak kelompok masyarakat sehingga kesalahpahaman yang ditemukan pada hasil *clustering* dapat berkurang.

5.2.4 Strategi 4 Kampanye “Kenali Bahan non-Halal”

Hasil analisis pada *Cluster 4* dan *Cluster 5* menunjukkan masih rendahnya pemahaman masyarakat mengenai bahan-bahan kritis halal yang dapat memengaruhi status kehalalan suatu produk. Sebgai masyarakat menganggap bahwa bahan tambahan yang digunakan dalam proses produksi juga perlu diperhatikan.

Padahal, beberapa bahan seperti *gelatin*, *emulsifer*, *enzim*, *shortening*, dan *flavor* (perasa) tertentu juga memerlukan verifikasi lebih lanjut untuk memastikan sumber dan proses produksinya sesuai dengan ketentuan halal. Namun, informasi mengenai bahan-bahan tersebut belum banyak dipahami oleh masyarakat sehingga menimbulkan kesenjangan pengetahuan mengenai pentingnya proses sertifikasi halal.

Untuk mengatasi kesenjangan tersebut, diusulkan kampanye edukasi bertema “Kenali Bahan Non-Halal sebelum Membeli” yang disampaikan melalui infografis di instastory, *Question an Answer* interaktif di instagram, dan konten seperti challenge di media sosial. Kampanye ini berisi penjelasan bahan-bahan kritis halal, alasan mengapa bahan perlu diperiksa, serta kaitannya dengan proses sertifikasi halal.



Gambar 5. 7 Contoh Implementasi Program Strategi 4

Gambar 5. 7 tersebut menampilkan informasi mengenai bebrbagai bahan yang perlu diperhatikan dalam proses sertifikasi halalm seperti *gelatin*, *emulsifer*, *enzim*, *shortening*, dan *falvor* (perasa). Selain itu, menamilkan alasan mengapa setiap bahan perlu diverifikasi untuk memastikan status halal nya.

Strategi ini didukung oleh penelitian di bidang literasi konsumen yang menunjukkan bahwa penyampaian informasi secara visual dan sederhana dapat meningkatkan pemahaman masyarakat seperti studi literatur yang dilakukan oleh (Q. Azizah & Dzikrulloh, 2024) dan (Putro et al., 2025) dari kedua penelitian ini menunjukkan bahwa program kampanye atau gerakan efektif untuk meningkatkan pemahaman tentang informasi yang disampaikan contoh dalam penelitian ini adalah kampanye dengan gerakan GEMESH (Gerakan Menuju Sertifikasi Halal). Melalui strategi ini, masyarakat diharapkan memiliki pemahaman yang lebih baik mengenai pentingnya pemeriksaan bahan baku dan bahan tambahan dalam proses sertifikasi halal sehingga dapat terus meningkatkan kesadaran untuk menerapkan sertifikasi halal.

5.3 Limitasi Penelitian

Pada penelitian ini tentu masih masih memiliki beberapa keterbatasan dalam proses nya. Diantaranya, yang pertama dalam proses pengolahan data pada penelitian ini menggunakan metode algoritma *Lexicon-based* untuk mengolah 961 data setelah *pre-processing*, namun metode ini memiliki kelauman yang bergantung pada sifat alami

metode ini yaitu masih bergantung pada kamus kata statis bahasa Indonesia, sehingga tidak semua dapat ditangkap dalam konteks yang lengkap. Dari hal ini membutuhkan validasi manual dalam menentukan sentimen nya. Lalu kedua, pada penelitian ini juga masih memiliki keterbatasan terhadap perolehan data yang hanya dari dua konten video di platform TikTok, sehingga generalisasi terhadap UMKM nasional kurang. Dan ketiga, penelitian ini masih memiliki keterbatasan dengan menggunakan validasi studi literatur, karena keterbatasan waktu penelitian hingga mendapat validasi menggunakan studi literatur saja.

BAB VI PENUTUP

6.1 Kesimpulan

Berikut merupakan kesimpulan dari penelitian yang telah dilakukan:

1. Pengelompokan komentar konsumen Indonesia terhadap kewajiban sertifikasi halal di platform media sosial TikTok menggunakan pendekatan *lexicon-based analysis* dan *clustering* ini menghasilkan *cluster* atau kelompok berjumlah lima *cluster* dengan total data sebesar 961 komentar, diantaranya terdapat *cluster* 1 dengan jumlah 546 data komentar dengan presentase 56,82%, lalu terdapat *cluster* 2 dengan jumlah 58 data komentar dengan presentase 6,04%. Kemudian *cluster* 3 dengan jumlah 197 data komentar. Selanjutnya, *cluster* 4 dengan jumlah 48 data komentar dengan presentase 4,99%. Dan terakhir, *cluster* 5 berjumlah 112 data komentar dengan presentase 11,65%.
2. Label dan proporsi sentimen yang diperoleh dari lima *cluster* yang telah terbentuk, ada *cluster* dengan sentimen negatif dengan proporsi sebanyak 693 data komentar dan presentase 72,11%, lalu sentimen positif sebanyak 161 data komentar dan presentase 16,75%, dan sentimen netral sebanyak 107 data komentar dan presentase 11,13%. Didapatkan hasil label masing-masing *cluster* diantaranya yang pertama, terdapat *cluster* 1 dengan sentimen negatif dan label *Profit-Oriented Perception* atau persepsi keuntungan ekonomi, lalu *cluster* 2 dengan sentimen negatif dan label *Cost Resistance* atau resistensi biaya, kemudian ada *cluster* 3 dengan sentimen positif dan label *Halal Awareness* atau kesadaran halal, selanjutnya *cluster* 4 dengan sentimen netral dan label *Regulation Confusion* atau kebingungan kebijakan, dan terakhir *cluster* 5 dengan sentimen negatif dan label *Halal Concern* atau kekhawatiran terhadap kehalalan produk.
3. Desain usulan strategi program percepatan sertifikasi halal diantaranya terdapat strategi 1 yaitu sosialisasi transparansi biaya menggunakan alat berbasis teknologi informasi (kalkulator BPJPH), yang diterapkan untuk *cluster* 1 dan *cluster* 2. Selanjutnya, strategi 2 yaitu kampanye testimoni UMKM yang berhasil meningkatkan penjualan setelah sertifikasi halal, yang diterapkan untuk *cluster* 2 dan 4. Lalu strategi 3 yaitu kolaborasi pemerintah, influencer, dan pelaku usaha halal dalam meningkatkan literasi digital sertifikasi halal melalui konten edukatif TikTok,

yang diterapkan untuk *cluster* 1, 2, 4, dan 5. Strategi terakhir yaitu strategi 4 Program Kampanye “*Kenali Bahan non-Halal*” untuk pemahaman masyarakat, yang diterapkan untuk *cluster* 4 dan *cluster* 5. Usulan strategi paling direkomendasikan yaitu strategi ke-3 karena strategi ini dapat diterapkan pada semua *cluster*.

6.2 Saran

Berikut merupakan saran atau rekomendasi pengembangan untuk penelitian selanjutnya:

1. Untuk penelitian selanjutnya disarankan peneliti memperoleh data dari media sosial TikTok, tetapi platform lain seperti media sosial Thread, Facebook, dan lainnya.
2. Untuk penelitian selanjutnya disarankan menggunakan kombinasi metode atau algoritma lain selain *Lexicon Based*, seperti *support vector machine*, *naïve bayes*, dan lainnya.
3. Pada penelitian selanjutnya, dalam waktu penelitian lebih lama, validasi hasil penelitian disarankan untuk menggunakan validasi secara survei kuantitatif atau Delphi dengan teknik konsensus ahli.

DAFTAR PUSTAKA

- Abidin, Z., & Niskaromah. (2025). The Influence of Social Media on Public Awareness of Halal Certification. *Zabags International Journal of Engagement*, 3(1), 35–43. <https://doi.org/10.61233/zijen.v3i1.32>
- Afandi, I. R., Fathurrohman, S., Wahyu, M., & Wibowo, A. (2025). Sentiment Analysis of Mandatory Halal Certification Policy on Twitter Using the Naive Bayes and K-Nearest Neighbors Algorithm. In *Halal Research* (Vol. 5, Number 1).
- Afrad, M., Febrianto, C., Wijayanto, S., & Fathoni, Y. (2024). Sentiment Analysis of Visitor Reviews on Baturaden Tourist Attraction Using Machine Learning Methods. *Edu Komputika*, 11(1). <https://journal.unnes.ac.id/journals/edukom/>
- Agusia, P., Uli, M., Manurung, A., Calista, V., & Mawardi, V. C. (2024). *SEMINAR NASIONAL CORISINDO Pemanfaatan WordCloud Pada Analisis Sentimen Dalam Menggali Persepsi Publik*.
- Arif Efendi, Fifi Hasmawati, & Syendi Apriko. (2025). Analisis Penggunaan Media TikTok Untuk Meningkatkan Kepedulian Lingkungan Siswa SMAN 1 Air Salek (Studi Kasus Pada Akun Tik Tok @Pandawaragroup). *Interaction Communication Studies Journal*, 1(4), 13. <https://doi.org/10.47134/interaction.v1i4.3811>
- Arif, M., Khamal Rokan, M., & Kumala, R. (2023). QUADRUPLE HELIX MODEL IN THE DEVELOPMENT OF HALAL MICRO BUSINESS IN NORTH SUMATRA. *Share: Jurnal Ekonomi Dan Keuangan Islam*, 12(1)(1), 201–224. <https://doi.org/10.22373/share.v12i1.15628>
- Astiwaru, E. M. (2025). ANALISIS DAMPAK KEBIJAKAN SERTIFIKASI HALAL TERHADAP PERTUMBUHAN INDUSTRI PANGAN. *JURNAL PAPATUNG*, 8(1).
- Azizah, N., Khaddapi, M., & Aqsa, M. (2025). *The Influence of Customer Testimonials and Product Quality on Purchasing Decisions in MSMEs in Pancasila Field, Palopo City*.
- Azizah, Q., & Dzikrulloh, D. (2024). Pengaruh Gerakan Sadar Halal dan Pengetahuan Produk Pada Gaya Hidup Halal Terhadap Keputusan Pembelian Gen Z. *Ekonomi, Keuangan, Investasi Dan Syariah (EKUITAS)*, 5(4), 645–654. <https://doi.org/10.47065/ekuitas.v5i4.5207>
- Bakti, I. T., & Ratnasih, C. (2025). Customer Perceptions of Digital Halal Certification and Its Implications for MSME Product Purchasing Decisions. *Greenation Internasional Journal of Economics and Accounting*, 3(4), 1–8.
- Chowdhury, S., & Alzarrad, A. (2023). Applications of *Text Mining* in the Transportation Infrastructure Sector: A Review. In *Information (Switzerland)* (Vol. 14, Number 4). MDPI. <https://doi.org/10.3390/info14040201>
- CNN Indonesia. (2019). Kantor Berita CNN. *CNN Indonesia*. <https://www.cnnindonesia.com/ekonomi/20191016145608-92-440050/mulai-besok-semua-produk-wajib-bersertifikat-halal> [01/02/2026]

- Ding, X., & Liu, B. (2010). *Resolving Object and Attribute Coreference in Opinion Mining*. <http://nlp.stanford.edu/software/tagger.shtml>
- Ding, X., Liu, B., & Yu, P. S. (2008). A holistic *lexicon-based* approach to opinion mining. *WSDM'08 - Proceedings of the 2008 International Conference on Web Search and Data Mining*, 231–239. <https://doi.org/10.1145/1341531.1341561>
- Erlangga, D., & Patimbangi, A. (2025). Analisis Minat Pelaku Umkm dalam Menggunakan Sertifikat Halal pada Kecamatan Tanete Riattang Barat Kabupaten Bone. *Ikraith Ekonomika*. <https://doi.org/10.37817/ikraith-ekonomika.v8i3>
- Fathoni, M. A. (2021). Potret Industri Halal Indonesia: Peluang dan Tantangan. *Jurnal Ilmiah Ekonomi Islam*, 6(3), 428. <https://doi.org/10.29040/jiei.v6i3.1146>
- Hakim, B. A. H., Mujahidah, A. S., & Rusydiana, A. S. (2022). SENTIMENT ANALYSIS ON HALAL CERTIFICATION. *Harmoni*, 21(1), 78–93. <https://doi.org/10.32488/harmoni.v21i1.609>
- Hardimanto, Z. Z. (2025). *The Appeal of Halal Labels: How much Influence Does It Have on Purchasing Decisions?* (Vol. 11, Number 3).
- Hasan, K. S. (2014). *KEPASTIAN HUKUM SERTIFIKASI DAN LABELISASI HALAL PRODUK PANGAN* ♣.
- Isnaina Akbari, Y., Safitri, D., & Sujarwo. (2025). *JICN: Jurnal Intelek dan Cendikiawan Nusantara UTILIZATION OF THE TIKTOK APLICATION AS A SOURCE OF INFORMATION FOR GEN Z*. <https://jicnusantara.com/index.php/jicn>
- Izza, N. N., Sari, M., & Shanty, M. K. (2022). Sentiment analysis of halal food trends during COVID-19. *Indonesian Journal of Islamic Economics Research*, 4(2), 64–75. <https://doi.org/10.18326/ijier.v4i2.7869>
- Jayus, D., Salim, N., & Rifqi Ma'arif, M. (2024). Analisis Persepsi Publik di Media Sosial Terhadap Produk Halal Di Indonesia Menggunakan *Text Mining*. In *Indonesian Journal on Data Science* (Vol. 2, Number 2). <https://ejournal.unjaya.ac.id/index.php/ijds>
- Jollyta, D., Efendi, S., Zarlis, M., & Mawengkang, H. (2019). *Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index*.
- Kamela, H. (2026). Peran Teknologi Informasi dalam Memoderasi Pengaruh Transparansi dan Akuntabilitas Biaya Pendidikan terhadap Persepsi Mahasiswa. *Jurnal Pendidikan Indonesia*, 6(2). <https://doi.org/10.62491/njpi.2026.v6i2-19>
- Kasturi, K., Sutisna, E., & Sekarini, R. A. (2025). Pengaruh Media Sosial Dan Sertifikasi Halal Terhadap Minat Beli Makanan Halal Mahasiswa Fakultas Ekonomi Universitas Islam Jakarta Pada Platform TikTok. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 7641–7648. <https://doi.org/10.31004/riggs.v4i2.1915>
- Kumar, A., & Sebastian, T. M. (2012). Sentiment Analysis: A Perspective on its Past, Present and Future. *International Journal of Intelligent Systems and Applications*, 4(10), 1–14. <https://doi.org/10.5815/ijisa.2012.10.01>

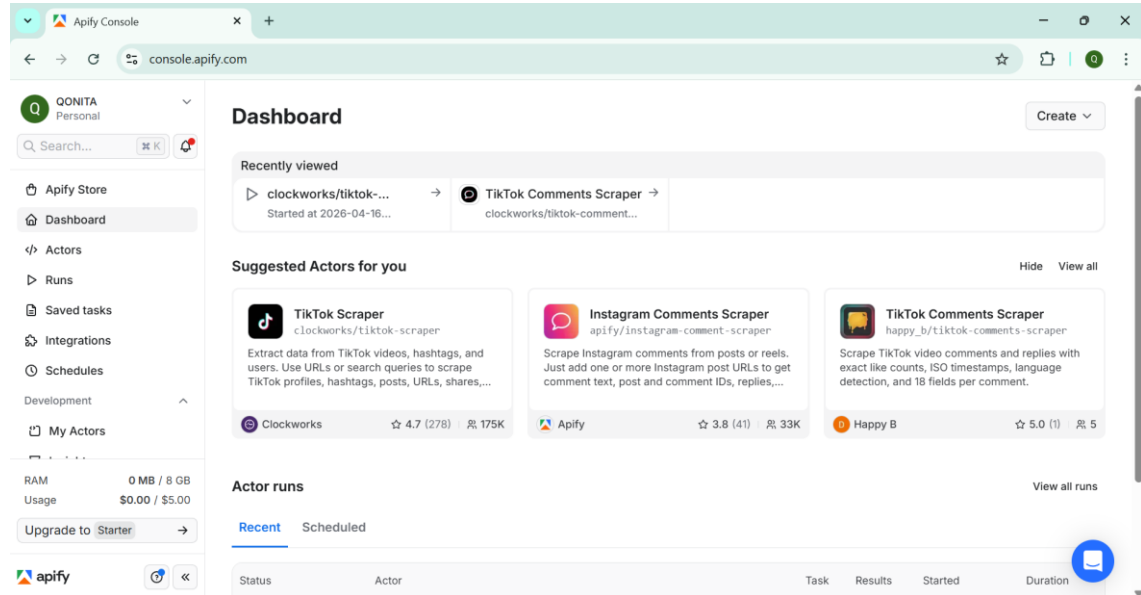
- Kumari, M., Jain, A., & Bhatia, A. (2016). Synonyms Based Term Weighting Scheme: An Extension to TF.IDF. *Procedia Computer Science*, 89, 555–561. <https://doi.org/10.1016/j.procs.2016.06.093>
- Larasati, A., Fadhilah Hayya, A., Az-Zakkiy, H., Seme, B. P., Jarot, H., & Hananto, D. (2023). Kajian Literatur: Dampak Sertifikasi Halal untuk Meningkatkan Daya Saing UMKM di Indonesia. In *Journal Of Global Humanistic Studies philosophiamundi.id* (Vol. 3, Number 1).
- Liu, B. (2020). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. <https://doi.org/DOI:10.1017/9781108639286>
- LPPOM MUI. (2008). *PANDUAN UMUM SISTEM JAMINAN HALAL LPPOM-MUI*. [LPPOM MUI] Lembaga Pengkajian Pangan, Obat-obatan, dan Kosmetik Majelis Ulama Indonesia. Panduan Umum Sistem Jaminan Halal LPPOM-MUI 2008. <http://www.halalmui.org/images/stories/pdf/sjh-indonesia.pdf>. [01/02/2026].
- Malimbe, A., Waani, F., & Suwu, E. A. A. (2021). *Dampak Penggunaan Aplikasi Online Tiktok (Douyin) Terhadap Minat Belajar di Kalangan Mahasiswa Sosiologi Fakultas Ilmu Sosial Dan Politik* (Vol. 1, Number 1).
- Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. In *Journal of King Saud University - Computer and Information Sciences* (Vol. 36, Number 4). King Saud bin Abdulaziz University. <https://doi.org/10.1016/j.jksuci.2024.102048>
- Maori, N. A. (2023). METODE ELBOW DALAM OPTIMASI JUMLAH *CLUSTER* PADA K-MEANS *CLUSTERING*. *Jurnal SIMETRIS*, 14.
- Mara, N., & Intisari, N. S. (2013). PENGKLASIFIKASIAN KARAKTERISTIK DENGAN METODE K-MEANS *CLUSTER ANALYSIS*. In *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)* (Vol. 02, Number 2).
- Masrukhan, M., & Ifrizah, ; (2024). *AI For Consumer Sentiment Analysis Towards Halal Products: A Study on Social Media Data* (Vol. 1, Number 1). <https://prosiding.areai.or.id/index.php/ICEAT>
- Muntholip, A., Setiawan, N., & Tinggi Ekonomi dan Bisnis Islam Al Rosyid Bojonegoro, S. (2025a). *Kajian Literatur: Dampak Sertifikasi Halal untuk Meningkatkan Daya Saing UMKM di Indonesia*. <https://jurnal.itscience.org/index.php/jumansi>
- Muntholip, A., Setiawan, N., & Tinggi Ekonomi dan Bisnis Islam Al Rosyid Bojonegoro, S. (2025b). *Sertifikasi Halal dan Daya Saing UMKM di Indonesia: Studi Systematic Literature Review*. <https://jurnal.itscience.org/index.php/jumansi>
- Nanta Setia Dien Alfansyuri, S., & Arsyi, N. (2023). THE EFFECT OF HALAL LABEL ON PRODUCT PACKAGING ON MUSLIM CONSUMER PURCHASE INTENTION. *Sensors*, 22. <https://doi.org/10.3390/xxxxx>
- Ningsih, R., Nadzary, V., Jannah, N., Rohman, K., Dzul, A., Issudi, C., Ghazani, H., & Nadzary A, V. (2025). *Pengaruh Influencer Muslim Terhadap Minat Beli Masyarakat Pada Produk Halal di E-Commerce*. 1(3). <https://doi.org/10.61166/abqori.v1i3.32>

- Nurwijayanto, P. R., Audia, S., Setyawati, D., Galih, A., & Sonalitha, E. (2024). AN EXAMINATION OF THE IMPACT OF PRICE LEVEL AND HALAL CERTIFICATION ON PURCHASE DECISION OF “ANGKRINGAN FOOD” MSME’s. In *Airlangga International Journal of Islamic Economics and Finance* (Vol. 7, Number 02). <https://doi.org/https://creativecommons.org/licenses/by/4.0/>
- Nury, A. F., Putri, I. P., & Juddi, M. F. (2024). Analisis Sertifikasi Halal Mixue Melalui Media Sosial Twitter Pada Akun @CNNINDONESIA. In *Jurnal Ilmiah Komunikasi* (Vol. 16, Number 02).
- Panayitsa, A. L., & Al Falah, P. (2024). Dampak Aplikasi Tiktok Terhadap Kebudayaan Indonesia. *Filosofi : Publikasi Ilmu Komunikasi, Desain, Seni Budaya*, 2(1), 75–88. <https://doi.org/10.62383/filosofi.v2i1.468>
- Parapat, D. A., & Azhar, A. A. (2024). Pemanfaatan Media Sosial Tiktok Sebagai Sumber Informasi Berita Bagi Mahasiswa FIS UIN Sumatera Utara Medan. *Oktober, Volume 2; Nomor 10*, 46–56. <https://doi.org/10.59435/gjmi.v2i10.943>
- Patel, P. D. (2024). Positive Online Customer Reviews Significantly Boost Sales for Micro-Businesses. *Integrated Journal for Research in Arts and Humanities*, 4(4), 85–92. <https://doi.org/10.55544/ijrah.4.4.14>
- Putri, W. T. H., & Hendrowati, R. (2018). PENGALIAN TEKS DENGAN MODEL BAG OF WORDS TERHADAP DATA TWITTER. 2(1), 129–138.
- Putro, H. S., Ariati, N. R. D., Gading, A. P., Ayundra, R., Salsabila, F. A., Hibatullah, F., Putri, A. M., Putra, A. S. E., Nawfal, M. F. J., Putra, M. F. L., Ciptoaji, Y. R., Khulaida, K., Purnomo, A. S., Putra, S. R., Rizqi, H. D., Wilujeng, S. A., Indrawan, I. A., & Tatas, T. (2025). Penerapan Kampanye Gerakan Menuju Sertifikasi Halal (GEMESH) Sebagai Metode Promosi dan Penjaminan Kualitas Perkopian Lokal di Area Kavling DPR Sidoarjo. *Sewagati*, 9(3), 726–735. <https://doi.org/10.12962/j26139960.v9i3.2589>
- Rahel, M., Murtasiah, M., & Salsabilaturribi, S. (2025). Signifikansi Sertifikasi Halal Yang Mengalahkan Promosi Dalam Keputusan Pembelian Konsumen Muslim Di TikTok Shop. *Iqtisad: Journal of Islamic Economic and Civilization*, 1(2), 222–235. <https://doi.org/10.61630/ijiec.v1i2.12>
- Rahmana, P. N., Putri N, D. A., & Damariswara, R. (2022). PEMANFAATAN APLIKASI TIKTOK SEBAGAI MEDIA EDUKASI DI ERA GENERASI Z. *Akademika*, 11(02), 401–410. <https://doi.org/10.34005/akademika.v11i02.1959>
- Rai, P., Singh, S., & Prof, A. (2010). A Survey of Clustering Techniques. In *International Journal of Computer Applications* (Vol. 7, Number 12).
- Razali, M. R., & Samori, Z. (2025). Exploring The Lived Experiences of Social Media Influencers in Shaping Halal Consumer Behavior Through Digital Engagement. *Journal of Halal Science and Management Research*, 1(1), 216–222. <https://doi.org/10.24191/jhsmr.v1i1.8995>
- Roesmara Donna, D., Danar Dono, N., & Iqbal Ahnaf, M. (2025). Evaluasi Dampak Sertifikasi Halal pada UMKM Produsen Makanan dan Minuman Anggota Desa Preneur Model K45PAK. *Indonesian Journal of Halal*, 8 (1)(1), 25–38. <https://doi.org/10.14710/halal.v8i1.25367>

- Sari, M., & Rahayu, S. S. (2024). *Sentiment Analysis on Halal Traceability: Dimensions-based Literature Index*. <http://journals.smartinsight.id/index.php/HS>
- Septiani, D., & Isabela, I. (2022). ANALISIS TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DALAM TEMU KEMBALI INFORMASI PADA DOKUMEN TEKS. *SINTESA: Jurnal Sistem Teknologi Dan Informasi Indonesia*, Vol. 1 No. 2, 81–87.
- Syahrani, H. (2025). Strategi Visual Tiktok Untuk Untuk Edukasi Generasi Z. In *Journal of Innovative and Creativity* (Vol. 5, Number 3).
- Timur, Y. P., Ratnasari, R. T., Pitchay, A. A., & Jamilu, U. (2023). Investigating Netizen Sentiment Toward Halal Certification in Indonesia Using Machine Learning. *Jurnal Ekonomi Syariah Teori Dan Terapan*, 10(6), 525–540. <https://doi.org/10.20473/vol10iss20236pp525-540>
- Triayudi, A., Pinastawa, I. W. R., & Pratama, J. D. (2024). *Clustering: Vol.1* (Edisi 1). PT Penamudamedia.
- Wani, A. A. (2024). Comprehensive analysis of *clustering* algorithms: exploring limitations and innovative solutions. *PeerJ Computer Science*, 10, 1–45. <https://doi.org/10.7717/PEERJ-CS.2286>
- Widaningsih, S., & Suheri, A. (2018). KLASIFIKASI JURNAL ILMU KOMPUTER BERDASARKAN PEMBAGIAN WEB OF SCIENCE DENGAN MENGGUNAKAN TEXT MINING. In *Seminar Nasional Teknologi Informasi dan Komunikasi*. www.elsevier.com,
- Yan, B., Wang, X., & Wu, J. (2020). Fiscal Transparency Online and Public Trust: An Exploratory Study on Baimiao Township Government. In *China Review* (Vol. 20, Number 3).

LAMPIRAN

Lampiran 1.1 *Scrapping Data*



Lampiran 1.2 *Programming (Google Colab)*

```
[x]
✓ 20s
!pip install sastrawi
!pip install nltk
!pip install swifter
!pip install networkx

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 7.0 MB/s eta 0:00:00
Installing collected packages: sastrawi
Successfully installed sastrawi-1.0.1
Requirement already satisfied: nltk in /usr/local/lib/python3.12/dist-packages (3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.12/dist-packages (from nltk) (8.3.3)
Requirement already satisfied: joblib in /usr/local/lib/python3.12/dist-packages (from nltk) (1.5.3)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.12/dist-packages (from nltk) (2025.11.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-packages (from nltk) (4.67.3)
Collecting swifter
  Downloading swifter-1.4.0.tar.gz (1.2 MB)
    1.2/1.2 MB 23.2 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: pandas>=1.0.0 in /usr/local/lib/python3.12/dist-packages (from swifter) (2.2.2)
Requirement already satisfied: psutil>=5.6.6 in /usr/local/lib/python3.12/dist-packages (from swifter) (5.9.5)
Requirement already satisfied: dask>=2.10.0 in /usr/local/lib/python3.12/dist-packages (from dask[dataframe]>=2.10.0->swifter) (2026.3.0)
Requirement already satisfied: tqdm>=4.33.0 in /usr/local/lib/python3.12/dist-packages (from swifter) (4.67.3)
Requirement already satisfied: click>=8.1 in /usr/local/lib/python3.12/dist-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (8.3.3)
Requirement already satisfied: cloudpickle>=3.0.0 in /usr/local/lib/python3.12/dist-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (3.1.2)
Requirement already satisfied: fsspec>=2021.09.0 in /usr/local/lib/python3.12/dist-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (2025.3.0)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (26.1)
Requirement already satisfied: partd>=1.4.0 in /usr/local/lib/python3.12/dist-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (1.4.2)
Requirement already satisfied: pyyaml>=5.3.1 in /usr/local/lib/python3.12/dist-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (6.0.2)
Requirement already satisfied: toolz>=0.12.0 in /usr/local/lib/python3.12/dist-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (0.12.1)
Requirement already satisfied: pyarrow>=16.0.0 in /usr/local/lib/python3.12/dist-packages (from dask[dataframe]>=2.10.0->swifter) (18.1.0)
Requirement already satisfied: numpy>=1.26.0 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.0.0->swifter) (2.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.0.0->swifter) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.0.0->swifter) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.0.0->swifter) (2026.1)
Requirement already satisfied: locket in /usr/local/lib/python3.12/dist-packages (from partd>=1.4.0->dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (1.0.0)
Requirement already satisfied: sio>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2->pandas>=1.0.0->swifter) (1.17.0)
Building wheels for collected packages: swifter
  Building wheel for swifter (setup.py) ... done
  Created wheel for swifter: filename=swifter-1.4.0-py3-none-any.whl size=16505 sha256=C382d21dca0f9326cbe38db025688dd0f1819825b727254fc0b25bd4e861439f1
  Stored in directory: /root/.cache/pip/wheels/d9/31/ff/ff51141a088571a9f672449e5a05e8b035c5d950a135f30
Successfully built swifter
Installing collected packages: swifter
Successfully installed swifter-1.4.0
ERROR: unknown command "instal" - maybe you meant "install"
```

import library

```

import pandas as pd
import numpy as np
import string
import re # regex library
import swifter
import networkx as nx

# NLTK
import nltk
from nltk.tokenize import word_tokenize
from nltk.tokenize.punkt import PunktSentenceTokenizer
from nltk.corpus import stopwords
from nltk.probability import FreqDist

# Sastrawi
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

# Sklearn
from sklearn.feature_extraction.text import TfidfTransformer, CountVectorizer

# Download resource NLTK
nltk.download('popular')

```

```

[nltk_data] Downloading collection 'popular'
[nltk_data] |
[nltk_data] | Downloading package cmudict to /root/nltk_data...
[nltk_data] | Unzipping corpora/cmudict.zip.
[nltk_data] | Downloading package gazetteers to /root/nltk_data...

```

```

data = pd.read_csv('/content/TA_Dataset_Sumber Tiktok.csv', on_bad_lines="warn")
data.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1099 entries, 0 to 1098
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  ---
0    text                  1099 non-null  object
1    diggCount             1099 non-null  int64
2    replyCommentTotal     1099 non-null  int64
3    createTimeISO         1099 non-null  object
4    uniqueId              1099 non-null  object
5    videoWebUrl           1099 non-null  object
6    uid                   1099 non-null  float64
7    cid                   1099 non-null  float64
8    avatarThumbnail       1099 non-null  object
dtypes: float64(2), int64(2), object(5)
memory usage: 77.4+ KB

```

```

df = pd.DataFrame(data[['createTimeISO', 'uniqueId', 'text']])
df.head(10)

```

	createTimeISO	uniqueId	text
0	2024-02-02T14:46:39.000Z	callmemoriz	Yang dagang babi keliling gimana? 🐷
1	2024-02-02T16:25:59.000Z	eddylin168	Beli bahannya mesti sama supplier yg ditunjuk ...
2	2024-02-02T05:05:22.000Z	nurshina24	buat beli rubikon lgi
3	2024-02-02T19:01:02.000Z	didiksupriyanto64	cuma jualan onde2 kira2 harus urus gak ya...ka...
4	2024-02-02T09:16:06.000Z	cina.toko	lebih penting setirikat halal apa sertifikat b...
5	2024-02-02T05:08:37.000Z	rizki.prtmaaa	ujung ujungnya bayar juga 🤔🤔
6	2024-02-02T09:05:18.000Z	ayppppppss	ini mah bisnis kalo kek gitu
7	2024-02-03T02:16:39.000Z	theo25264	pajak berkedok sertifikat 🐷
8	2024-02-02T16:38:40.000Z	esteh1993	Moga aja gak ribet urusnya + gratis gak pake b...
9	2024-02-02T05:33:53.000Z	nugisuryaperdana	mending bikin stiker sendiri

Next steps: [Generate code with df](#) [New interactive sheet](#)

```
[3]
✓ On
df = pd.read_csv('/content/TA_Dataset_Sumber Tiktok.csv')
df.head()
```

	text	diggCount	replyCommentTotal	createTimeISO	uniqueId	videoUrl	uid	cid	avatarThumbnail
0	Yang dagang babi kelling gimana? 🐷	13	5	2024-02-02T14:46:39.000Z	calmemoriz	https://www.tiktok.com/@geraldvincent/video/7...	6.940000e+18	7.330000e+18	https://p19-common-sign.tiktokcdn-us.com/tos-m...
1	Bel bahannya mesti sama supplier yg ditunjuk ...	1	0	2024-02-02T16:25:59.000Z	eddylin168	https://www.tiktok.com/@geraldvincent/video/7...	6.500000e+18	7.330000e+18	https://p16-common-sign.tiktokcdn-us.com/tos-m...
2	buat beli rubikon lgi	313	7	2024-02-02T05:05:22.000Z	nurshina24	https://www.tiktok.com/@geraldvincent/video/7...	7.230000e+18	7.330000e+18	https://p16-common-sign.tiktokcdn-us.com/tos-m...
3	cuma jualan onde2 kira2 harus unus gak ya ka...	7	0	2024-02-02T19:01:02.000Z	didksupriyanto64	https://www.tiktok.com/@geraldvincent/video/7...	6.990000e+18	7.330000e+18	https://p16-common-sign.tiktokcdn-us.com/tos-u...
4	lebih penting setingkat halai apa sertifikat b...	77	31	2024-02-02T09:16:06.000Z	cina toko	https://www.tiktok.com/@geraldvincent/video/7...	6.780000e+18	7.330000e+18	https://p16-common-sign.tiktokcdn-us.com/tos-m...

```
[4]
✓ On
df.info()
df.duplicated(subset='text').sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1099 entries, 0 to 1098
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype
---  ---
 0   text        1099 non-null   object
 1   diggCount   1099 non-null   int64
 2   replyCommentTotal  1099 non-null   int64
 3   createTimeISO  1099 non-null   object
 4   uniqueId     1099 non-null   object
 5   videoUrl     1099 non-null   object
 6   uid          1099 non-null   float64
 7   cid          1099 non-null   float64
 8   avatarThumbnail  1099 non-null   object
dtypes: float64(2), int64(2), object(5)
memory usage: 77.4+ MB
np.int64(33)
```

cleaning text

```
[7]
✓ On
# Load dataset
df = pd.read_csv("/content/TA_Dataset_Sumber Tiktok.csv")

# 1. REMOVE USERNAME (@username)
def remove_username(text):
    if text and isinstance(text, str):
        return re.sub(r'@[^\s]+', '', text)
    return text

# 2. REMOVE EMOJI
def remove_emoji(text):
    if text is None or not isinstance(text, str):
        return text

    emoji_pattern = re.compile(
        "[
            u'\U0001F600-\U0001F64F'
            u'\U0001F300-\U0001F5FF'
            u'\U0001F680-\U0001F6FF'
            u'\U0001F700-\U0001F77F'
            u'\U0001F780-\U0001F7FF'
            u'\U0001F800-\U0001F8FF'
            u'\U0001F900-\U0001F9FF'
            u'\U0001FA00-\U0001FAGF'
            u'\U0001FA70-\U0001FAFF'
            u'\U00002702-\U000027B0'
            u'\U000024C2-\U0001F251'
        ]+", flags=re.UNICODE
    )
    return emoji_pattern.sub(r'', text)
```

```

# 5. LOWERCASE
def to_lower(text):
    if text and isinstance(text, str):
        return text.lower()
    return text

# 6. REMOVE EXTRA SPACES
def remove_extra_spaces(text):
    if text and isinstance(text, str):
        return re.sub(r'\s+', ' ', text).strip()
    return text

# APPLY CLEANING PIPELINE (TEXT CLEANING)
df["cleaning"] = df["text"].apply(remove_username)
df["cleaning"] = df["cleaning"].apply(remove_emoji)
df["cleaning"] = df["cleaning"].apply(remove_symbols)
df["cleaning"] = df["cleaning"].apply(remove_numbers)
df["cleaning"] = df["cleaning"].apply(to_lower)
df["cleaning"] = df["cleaning"].apply(remove_extra_spaces)

# HASIL
df.head()

```

	text	diggCount	replyCommentTotal	createTimeISO	uniqueId
0	Yang dagang babi keliling gimana? 🐷	13	5	2024-02-02T14:46:39.000Z	callmemoriz h
1	Beli bahannya mesti sama supplier yg ditunjuk ...	1	0	2024-02-02T16:25:59.000Z	eddylin168 h
2	buat beli rubikon lgi	313	7	2024-02-02T05:05:22.000Z	nurshina24 h
3	cuma jualan onde2 kira2 harus urus gak ya...ka...	7	0	2024-02-02T19:01:02.000Z	didiksupriyanto64 h
4	lebih penting setirikat halal apa sertifikat b...	77	31	2024-02-02T09:16:06.000Z	cina.toko h

Next steps: [Generate code with df](#) [New interactive sheet](#)

Case Folding

```

[8]
✓ 0s
def case_folding(text):
    if isinstance(text, str):
        lowercase_text = text.lower()
        return lowercase_text
    return text

df['case_folding'] = df['cleaning'].apply(case_folding)
df.head(5)

```

	text	diggCount	replyCommentTotal	createTimeISO	
0	Yang dagang babi keliling gimana? 🐷	13	5	2024-02-02T14:46:39.000Z	
1	Beli bahannya mesti sama supplier yg ditunjuk ...	1	0	2024-02-02T16:25:59.000Z	
2	buat beli rubikon lgi	313	7	2024-02-02T05:05:22.000Z	
3	cuma jualan onde2 kira2 harus urus gak ya...ka...	7	0	2024-02-02T19:01:02.000Z	didik:
4	lebih penting setirikat halal apa sertifikat b...	77	31	2024-02-02T09:16:06.000Z	

Next steps: [Generate code with df](#) [New interactive sheet](#)

Normalisasi kata + Code

```

9]
+ Os
import pandas as pd

kamus = pd.read_csv("/content/slang.csv")
kamus_tidak_baku = dict(zip(kamus['slang'], kamus['formal']))

10]
+ Os
def replace_tidak_baku(text, kamus):
    words = text.split()

    hasil = []
    kata_baku = []
    kata_diganti = []

    for word in words:
        if word in kamus:
            hasil.append(kamus[word])
            kata_baku.append(kamus[word])
            kata_diganti.append(word)
        else:
            hasil.append(word)

    return " ".join(hasil), kata_baku, kata_diganti

11]
+ Os
df['normalisasi'] = df['case_folding'].apply(
    lambda x: replace_tidak_baku(x, kamus_tidak_baku)[0]
)

df['kata_baku'] = df['case_folding'].apply(
    lambda x: replace_tidak_baku(x, kamus_tidak_baku)[1]
)

df['kata_diganti'] = df['case_folding'].apply(
    lambda x: replace_tidak_baku(x, kamus_tidak_baku)[2]
)

df.head()

```

	text	diggCount	replyCommentTotal	createTimeISO	uniqueId	videoWebUrl
0	Yang dagang babi keliling gimana? 🐷	13	5	2024-02-02T14:46:39.000Z	callmemoriz	https://www.tiktok.com/@geraldvincentt/video/7... 6.94000
1	Beli bahannya mesti sama supplier yg ditunjuk ...	1	0	2024-02-02T16:25:59.000Z	eddylin168	https://www.tiktok.com/@geraldvincentt/video/7... 6.50000
2	buat beli rubikon lgi	313	7	2024-02-02T05:05:22.000Z	nurshina24	https://www.tiktok.com/@geraldvincentt/video/7... 7.23000
3	cuma jualan onde2 kira2 harus urus gak ya...ka...	7	0	2024-02-02T19:01:02.000Z	didiksupriyanto64	https://www.tiktok.com/@geraldvincentt/video/7... 6.99000
4	lebih penting setirikat halal apa sertifikat b...	77	31	2024-02-02T09:16:06.000Z	cina toko	https://www.tiktok.com/@geraldvincentt/video/7... 6.78000

Tokenize

```

12]
+ Os
def tokenize(text):
    return text.split()

df['tokens'] = df['normalisasi'].apply(tokenize)

# SPLIT DATA 80:20
from sklearn.model_selection import train_test_split

X = df['tokens']

X_train, X_test = train_test_split(
    X,
    test_size=0.2,
    random_state=42
)

df.head()

```

	text	diggCount	replyCommentTotal	createTimeISO	uniqueId	videoWebUrl
0	Yang dagang babi keliling gimana? 🐷	13	5	2024-02-02T14:46:39.000Z	callmemoriz	https://www.tiktok.com/@geraldvincentt/video/7... 6.94000
1	Beli bahannya mesti sama supplier yg ditunjuk ...	1	0	2024-02-02T16:25:59.000Z	eddylin168	https://www.tiktok.com/@geraldvincentt/video/7... 6.50000
2	buat beli rubikon lgi	313	7	2024-02-02T05:05:22.000Z	nurshina24	https://www.tiktok.com/@geraldvincentt/video/7... 7.23000
3	cuma jualan onde2 kira2 harus urus gak ya...ka...	7	0	2024-02-02T19:01:02.000Z	didiksupriyanto64	https://www.tiktok.com/@geraldvincentt/video/7... 6.99000
4	lebih penting setirikat halal apa sertifikat b...	77	31	2024-02-02T09:16:06.000Z	cina toko	https://www.tiktok.com/@geraldvincentt/video/7... 6.78000

Next steps: [Generate code with df](#) [New interactive sheet](#)

Stopword Removal

```
[13]
✓ 0s
import nltk
nltk.download('stopwords')

from nltk.corpus import stopwords
stop_words = set(stopwords.words("indonesian"))

list_stopwords = set([
    # kata umum
    "dan", "atau", "karena", "jadi", "agar", "supaya", "untuk", "dari", "ke", "di",
    "pada", "oleh", "sebagai", "dalam", "itu", "ini", "tersebut", "yang", "dengan", "juga",

    # filler
    "aja", "doang", "banget", "nih", "sih", "tuh", "dong", "deh", "kok", "loh", "lah",

    # slang
    "yg", "dg", "dgn", "ny", "krn", "tdk", "utk", "jd", "jgn", "sdh", "klo", "kalo", "lgi",

    # sosmed noise
    "wkwk", "wk", "haha", "hehe", "hihi", "amp", "amp;", "nbsp", "hasan"
])

def stopwords_removal(words):
    return [word for word in words if word not in list_stopwords]

df['stopword_removal'] = df['tokens'].apply(stopwords_removal)

df.head()
```

... [nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!

	text	diggCount	replyCommentTotal	createTimeISO	uniqueId	videoWebUrl
0	Yang dagang babi kelling gimana? 🐷	13	5	2024-02-02T14:46:39.000Z	callmemoriz	https://www.tiktok.com/@geraldvincentt/video/7...

Stemming

```
[14]
✓ 0s
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

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

[15]
✓ 0s
def stem_text(tokens):
    return [stemmer.stem(word) for word in tokens]

[16]
✓ 2m
df['stemming_data'] = df['stopword_removal'].apply(stem_text)
df.head()
```

	text	diggCount	replyCommentTotal	createTimeISO	uniqueId	videoWebUrl
0	Yang dagang babi kelling gimana? 🐷	13	5	2024-02-02T14:46:39.000Z	callmemoriz	https://www.tiktok.com/@geraldvincentt/video/7... 6.5
1	Beli bahannya mesti sama supplier yg ditunjuk ...	1	0	2024-02-02T16:25:59.000Z	eddylin168	https://www.tiktok.com/@geraldvincentt/video/7... 6.5
2	buat beli rubikon lgi	313	7	2024-02-02T05:05:22.000Z	nurshina24	https://www.tiktok.com/@geraldvincentt/video/7... 7.2
3	cuma jualan onde2 kira2 harus	7	0	2024-02-02T19:01:02.000Z	didiksupriyanto64	https://www.tiktok.com/@geraldvincentt/video/7... 6.5

TF-IDF

```
[18]
✓ Os  import pandas as pd
import ast

# data real
df_real = pd.read_csv('/content/TA_Dataset_Sumber Tiktok.csv')

# data bersih
df_clean = pd.read_csv('/content/TA_Data_Clean_betul.csv')

[19]
✓ Os
def to_text(x):
    try:
        if isinstance(x, str):
            x = ast.literal_eval(x)
            return ' '.join(x)
        except:
            return ''

df_clean['text_final'] = df_clean['stemming_data'].apply(to_text)
```

menghitung kata

```
20]
✓ Os from sklearn.feature_extraction.text import TfidfVectorizer

# Memastikan tidak ada NaN (data kosong)
df_clean['text_final'] = df_clean['text_final'].fillna('')

# TF-IDF
tfidf = TfidfVectorizer(max_features=1000)

X_tfidf = tfidf.fit_transform(df_clean['text_final'])

print(X_tfidf.shape)

(1099, 1000)

21]
✓ Os feature_names = tfidf.get_feature_names_out()

df_tfidf = pd.DataFrame(
    X_tfidf.toarray(),
    columns=feature_names
)

print(df_tfidf.head())

abg  abud  abunawas  abut      ada  adaaja  adalah  adat  adik  adil  ...  \
0  0.0  0.0      0.0  0.0  0.000000  0.0  0.0  0.0  0.0  0.0  ...
1  0.0  0.0      0.0  0.0  0.000000  0.0  0.0  0.0  0.0  0.0  ...
2  0.0  0.0      0.0  0.0  0.000000  0.0  0.0  0.0  0.0  0.0  ...
3  0.0  0.0      0.0  0.0  0.198892  0.0  0.0  0.0  0.0  0.0  ...
4  0.0  0.0      0.0  0.0  0.000000  0.0  0.0  0.0  0.0  0.0  ...

yahjhh  yai  yakin  yakni  yaman  yasudah  yau  yee  ygy  zaman
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
2  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
4  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0

[5 rows x 1000 columns]

22]
✓ Os # Menggabungkan text_final dengan hasil TF-IDF
df_hasil = pd.concat([df_clean[['text_final']], df_tfidf], axis=1)
```

PENGOLAHAN DATA

algoritma K-Means Clustering

```
[31]
✓ Os
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

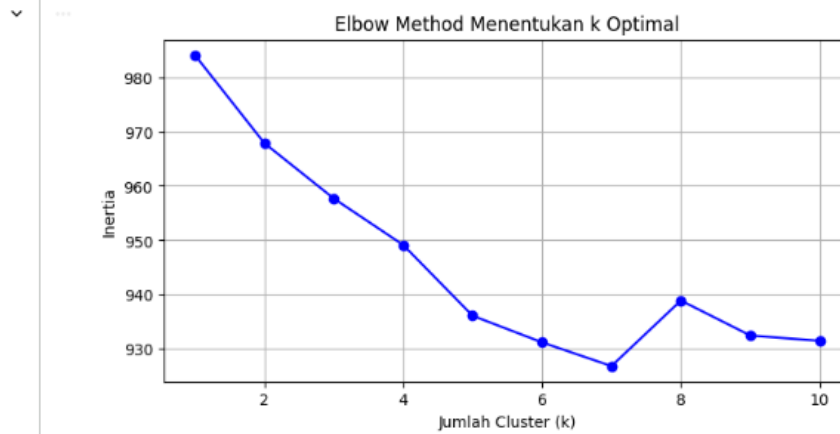
inertia = []
K = range(1, 11)

for k in K:
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(X_tfidf) # lgsg pake ini
    inertia.append(kmeans.inertia_)
```

```
[32]
✓ Os
plt.figure(figsize=(8, 4))
plt.plot(K, inertia, 'bo-')

plt.xlabel("Jumlah Cluster (k)")
plt.ylabel("Inertia")
plt.title("Elbow Method Menentukan k Optimal")

plt.grid(True)
plt.show()
```



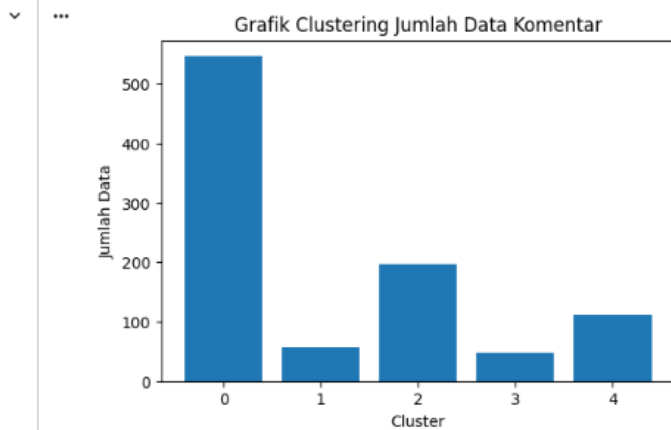
```
[48] import pandas as pd
✓ Os import matplotlib.pyplot as plt

# Menghitung jumlah data tiap cluster
cluster_counts = df['cluster'].value_counts().sort_index()

# Plot bar chart
plt.figure(figsize=(6,4))
plt.bar(cluster_counts.index, cluster_counts.values)

plt.xlabel('cluster')
plt.ylabel('Jumlah Data')
plt.title('Grafik Clustering Jumlah Data Komentar')

plt.xticks(cluster_counts.index)
plt.show()
```



lexicon based analysis

```
[49]
✓ 0s import pandas as pd

# Load hasil preprocessing
data = pd.read_csv("/content/TA_Data_Clean_TF-IDF.csv")
data.info()

# nb: hanya kolom stemming_data
data = pd.DataFrame(data["text_final"])
data.head(5)
```

```
... <class 'pandas.core.frame.DataFrame'>
RangeIndex: 1099 entries, 0 to 1098
Data columns (total 18 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   text                  1099 non-null  object
 1   diggCount             1099 non-null  int64
 2   replyCommentTotal     1099 non-null  int64
 3   createTimeISO         1099 non-null  object
 4   uniqueId              1099 non-null  object
 5   videoWebUrl           1099 non-null  object
 6   uid                   1099 non-null  float64
 7   cid                   1099 non-null  float64
 8   avatarThumbnail       1099 non-null  object
 9   cleaning              1032 non-null  object
10   case_folding          1032 non-null  object
11   normalisasi           1032 non-null  object
12   kata_baku             1099 non-null  object
13   kata_diganti          1099 non-null  object
14   tokens                1099 non-null  object
15   stopword_removal       1099 non-null  object
16   stemming_data         1099 non-null  object
17   text_final            1030 non-null  object
dtypes: float64(2), int64(2), object(14)
memory usage: 154.7+ KB
```

	text_final
0	dagang babi keliling gimana
1	beli bahan mesti sama supplier tunjuk apalagi ...
2	buat beli rubikon lagi
3	hanya jual onde kira harus urus tidak ya kan b...
4	lebih penting setirikat halal apa sertifikat b...

Cluster 1

```
[4]
✓ 2s import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS

df = pd.read_csv('/content/Cluster_1.csv', encoding='latin-1')
df.head(10)

# ambil teks dari kolom Cluster_1
text = " ".join(df['Cluster_1'].dropna().astype(str))

# stopwords
stopwords = set(STOPWORDS)
stopwords.update(["https", "aep", "ya", "aja", "nih", "yg", "nya"])

# buat wordcloud
wc = WordCloud(
    stopwords=stopwords,
    background_color="white",
    max_words=200,
    width=800,
    height=400,
    collocations=False
).generate(text)

# tampilkan
plt.figure(figsize=(10,5))
plt.imshow(wc)
plt.axis("off")
plt.title("WordCloud Cluster 1")
plt.show()
```

[illegible]

WordCloud Cluster 4



WordCloud Cluster 5



