

Analisis Kluster Judul Berita di Website Suarapublik.Id Menggunakan Metode K-Means

Tania Galuh Parwati¹, Rina Antasari^{2*}, Eriene Dheanda Absharina³

¹ Institut Teknologi dan Sains Nahdlatul Ulama Sriwijaya Sumsel, Program Studi Informatika, Institut Teknologi dan Sains, Palembang, Indonesia

Email: ¹taniaparwati39@gmail.com, ^{2*}rinaantasari_uin@radenfatah.ac.id, ^{3*}eriedheanda@itsnusriwijaya.ac.id
(* Email Corresponding Author: taniaparwati39@gmail.com)

Received: February 25, 2026 | Revision: February 26, 2026 | Accepted: February 27, 2026

Abstrak

Perkembangan media berita daring menyebabkan meningkatnya jumlah informasi yang dipublikasikan setiap hari, sehingga diperlukan metode analisis yang mampu mengelompokkan informasi secara sistematis. Website suarapublik.id sebagai media yang bergerak di bidang informasi dan komunikasi menghasilkan beragam judul berita dengan topik yang bervariasi, sehingga berpotensi menyulitkan dalam analisis pola pemberitaan. Penelitian ini bertujuan untuk mengelompokkan judul berita pada website suarapublik.id menggunakan metode K-Means Clustering berbasis text mining. Kajian pustaka dalam penelitian ini mengacu pada konsep text mining, pra-pemrosesan teks, pembobotan kata TF-IDF, serta algoritma K-Means sebagai metode klusterisasi data teks. Metode penelitian yang digunakan adalah kuantitatif dengan objek penelitian berupa 100 judul berita yang dikumpulkan secara manual dan diolah menggunakan platform Google Colab. Data dianalisis melalui tahapan pra-pemrosesan teks, pembobotan TF-IDF, penentuan jumlah kluster optimal menggunakan Metode Elbow, serta proses klusterisasi K-Means. Hasil penelitian menunjukkan terbentuknya tujuh kluster dengan tema pemberitaan yang berbeda, mencerminkan pola dan kecenderungan topik pada suarapublik.id. Penelitian ini membuktikan bahwa metode K-Means efektif dalam mengelompokkan judul berita secara otomatis dan sistematis.

Kata Kunci: TF-IDF, K-Means, klusterisasi berita, judul berita

Abstract

The rapid development of online news media has led to a significant increase in the amount of information published daily, requiring analytical methods capable of systematically organizing information. Suarapublik.id, as an online news platform operating in the field of information and communication, publishes diverse news headlines covering various topics, which may complicate the identification of content patterns. This study aims to cluster news headlines on the suarapublik.id website using the K-Means Clustering method based on text mining techniques. The literature review of this research is grounded in the concepts of text mining, text preprocessing, TF-IDF weighting, and the K-Means algorithm as a text clustering method. This research employs a quantitative approach with 100 news headlines collected manually and processed using the Google Colab platform. Data analysis was conducted through text preprocessing, TF-IDF weighting, determination of the optimal number of clusters using the Elbow Method, and K-Means clustering. The results show that seven distinct clusters were formed, each representing different news themes, reflecting the content patterns and topic tendencies on suarapublik.id. This study demonstrates that the K-Means method is effective in automatically and systematically clustering news headlines.

Keywords: text mining, TF-IDF, K-Means, news clustering, news headlines

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah membawa perubahan signifikan dalam pola konsumsi informasi masyarakat, khususnya melalui media berita online. Media digital memungkinkan penyebaran informasi secara cepat, luas, dan berkelanjutan, sehingga mendorong peningkatan produksi konten teks dalam jumlah besar setiap harinya. Judul berita sebagai bagian penting dari konten media online berfungsi sebagai representasi ringkas dari isi berita dan menjadi pintu utama bagi pembaca dalam memahami topik yang disajikan. Akumulasi judul berita tersebut membentuk data teks tidak terstruktur yang sulit dianalisis secara manual tanpa pendekatan komputasional yang memadai [1][2].

Meskipun jumlah data judul berita terus meningkat, pengelolaan dan pengelompokan judul berita berdasarkan tema masih menjadi permasalahan utama pada banyak portal berita online. Judul berita yang dipublikasikan tanpa pengelompokan tematik yang jelas menyulitkan proses identifikasi pola, kecenderungan topik, dan distribusi isu pemberitaan. Penelitian terdahulu menunjukkan bahwa data teks tidak terstruktur, seperti judul berita, tidak dapat dianalisis secara optimal menggunakan metode statistik konvensional, sehingga diperlukan teknik khusus berbasis text mining untuk mengekstraksi informasi yang bermakna [3][4].

Urgensi penelitian ini semakin meningkat seiring dengan kebutuhan media digital terhadap sistem pengelolaan konten yang lebih sistematis dan objektif. Penerapan teknik text mining tidak hanya relevan secara ilmiah dalam pengembangan analisis data teks, tetapi juga memiliki nilai praktis dalam membantu media online memahami kecenderungan pemberitaan, meningkatkan kualitas kategorisasi konten, serta mendukung pengambilan keputusan redaksional berbasis data. Oleh karena itu, penggunaan algoritma klusterisasi menjadi solusi yang potensial untuk mengelompokkan judul berita secara otomatis dan efisien.

Penelitian ini bertujuan untuk menganalisis pengelompokan judul berita pada website suarapublik.id menggunakan metode K-Means berbasis pembobotan kata Term Frequency–Inverse Document Frequency (TF-IDF). Metode ini dipilih karena mampu merepresentasikan karakteristik teks dalam bentuk vektor numerik dan mengelompokkan data berdasarkan tingkat kemiripan tanpa memerlukan data berlabel. Dengan demikian, diharapkan dapat dihasilkan kluster judul berita yang merepresentasikan tema pemberitaan secara jelas dan terstruktur.

Kontribusi artikel ini terletak pada penerapan metode text mining dan algoritma K-Means dalam konteks analisis judul berita media online lokal, khususnya suarapublik.id. Hasil penelitian ini diharapkan dapat memperkaya kajian ilmiah di bidang pengolahan teks dan data mining, serta memberikan referensi praktis bagi pengelola media digital dalam mengembangkan sistem pengelompokan dan analisis konten berita berbasis data.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksploratif dengan metode text mining. Pendekatan ini dipilih karena penelitian bertujuan mengolah data teks menjadi representasi numerik serta mengidentifikasi pola tema pemberitaan secara otomatis menggunakan algoritma klusterisasi.

Objek penelitian adalah media online lokal suarapublik.id. Data penelitian berupa judul berita yang dipublikasikan pada website tersebut dan dikumpulkan selama periode penelitian tahun 2025. Seluruh judul berita yang memenuhi kriteria kelengkapan data digunakan sebagai sampel penelitian dengan teknik total sampling, sehingga analisis dapat merepresentasikan pola pemberitaan secara menyeluruh.

Instrumen penelitian berupa perangkat lunak pengolahan data berbasis Python. Proses pengumpulan data dilakukan melalui dokumentasi digital, kemudian dilanjutkan dengan tahapan pra-pemrosesan teks yang meliputi case folding, tokenizing, dan stopword removal untuk memastikan keseragaman dan kebersihan data teks [5].

Analisis data dilakukan menggunakan metode pembobotan TF-IDF untuk merepresentasikan judul berita dalam bentuk vektor numerik. Selanjutnya, algoritma K-Means digunakan untuk mengelompokkan judul berita berdasarkan tingkat kemiripan kata. Penentuan jumlah kluster optimal dilakukan menggunakan Elbow Method, sedangkan interpretasi hasil kluster dilakukan dengan mengidentifikasi kata-kata dominan pada masing-masing kluster [6][7].

Table 1. Alur Metode Penelitian

Tahap	Metode	Output
Pengumpulan Data	Dokumentasi digital	Dataset judul berita
Pra-pemrosesan	Case folding, tokenizing, stopword removal	Data teks bersih
Pembobotan	TF-IDF	Vektor numerik
Klusterisasi	K-Means + Elbow Method	Kluster tema berita

3. HASIL DAN PEMBAHASAN

Analisis data dilakukan terhadap kumpulan judul berita pada website suarapublik.id yang telah melalui tahapan pra-pemrosesan teks berupa case folding, tokenizing, stopword removal, dan stemming.



index	Judul Berita	case_folding
0	Ekspor Santan Sumsel Tembus 543 Ton, China Jadi Pasar Utama	ekspor santan sumsel tembus 543 ton, china jadi pasar utama
1	Wanita Muda di Palembang Ditemukan Tewas di Kamar Hotel	wanita muda di palembang ditemukan tewas di kamar hotel
2	Peningkatan Akses Keuangan Daerah untuk Perkuat Program Prioritas Pemerintah	peningkatan akses keuangan daerah untuk memperkuat program prioritas pemerintah
3	Oknum Brimob di Ogan Ilir Dilaporkan Istri ke Propam, Ini Dugaan Kasusnya	oknum brimob di ogan ilir dilaporkan istri ke propam, ini dugaan kasusnya
4	Kasus Dugaan Asusila Oknum Dosen Terus Bergulir, Ini Tanggapan BEM KM FISIP Unsrri	kasus dugaan asusila oknum dosen terus bergulir, ini tanggapan bem km fisip unsrri
5	Pastikan Pasokan Listrik Aman Jelang Porprov–Peparprov Sumsel 2025 PLN ULP Sekayu Cek Sistem Jaringan	pastikan pasokan listrik aman jelang porprov–peparprov sumsel 2025 pln ulp sekayu cek sistem jaringan
6	Polsek BMT Gelar Patroli Rutin Antisipasi Aksi Balap Liar	polsek bmt gelar patroli rutin antisipasi aksi balap liar
7	Terceduk Hendak Mencuri, Junaidi Diamankan Warga Babatan Saudagar	terceduk hendak mencuri, junaidi diamankan warga babatan saudagar
8	Kegiatan Illegal Drilling di Wilayah Gunung Megang Dibongkar Polres Muara Enim	kegiatan ilegal drilling di wilayah gunung megang dibongkar polres muara enim
9	Kapolda Sumsel: Anggota Poli Terlibat Narkoba Bakal Dipecat!	kapolda sumsel. anggota poli terlibat narkoba bakal dipecat!
10	Uji Kemampuan dan Kualitas, 15 Mahasiswa Prodi PAI UIN Ikufti Ujian Munaqosyah	uji kemampuan dan kualitas, 15 mahasiswa prodi pai uin ikufti ujian munaqosyah
11	Motor Digelapkan Mantan Suami, Eva Lapor Polisi	motor digelapkan mantan suami, eva lapor polisi
12	Agar Proses Pembelajaran Berjalan Sesuai Kurikulum, Guru MAN 1 Disupervisi	agar proses pembelajaran berjalan sesuai kurikulum, guru man 1 disupervisi
13	PPPJKB Tolak Kenaikan Tarif Retribusi, Kembalikan Lapak Puluhan Pedang Pasar Jakabaring!	pppjkbtolak kenaikan tarif retribusi, kembalikan lapak puluhan pedang pasar jakabaring!
14	Tim Opsnal Macan Putih Polsek BMT Gelar Patroli Hunting Sasar Daerah Rawan	tim opsnal macan putih polsek bmt gelar patroli hunting sasar daerah rawan
15	Karhutla Hanguskan 2 Hektare Lahan di Jejawi OKI	karhutla hanguskan 2 hektare lahan di jejawi oki
16	Dua Terdakwa Kasus Korupsi DPMD Muba Dituntut 3 dan 4 Tahun Penjara	dua terdakwa kasus korupsi dpmd muba dituntut 3 dan 4 tahun penjara
17	Dewa Minta Hasil Program Pemberdayaan Warga Binaan Lapas Palembang di Salurkan Ke UMKM	dewa minta hasil program pemberdayaan warga binaan lapas palembang di salurkan ke umkm
18	Pastikan Jajanan Siswa Sesuai Standar Kebersihan Kepala Madrasah Sidak Kantin Sekolah	pastikan jajanan siswa sesuai standar kebersihan kepala madrasah sidak kantin sekolah
19	SMA 6 Gandeng TNI, Bentuk Karakter Siswa dan Cegah Kenakalan Remaja	sma 6 gandeng tni, bentuk karakter siswa dan cegah kenakalan remaja
20	Kejati Telapkan Dua Tersangka Kasus Jaksa Gadungan, Salah satunya PNS Way Kanan	kejati telapkan dua tersangka kasus jaksa gadungan, salah satunya pns way kanan
21	Hakim Sarankan Restorasi Justice Dalam Kasus Domeslesan Surat wasno Menteri Oknum Polici	hakim sarankan restorasi justice dalam kasus domeslesan surat wasno menteri oknum polici

Gambar 1. Fase Folding

Index	case_folding	tokenizing
0	ekspor santan surnel tembus 543 ton, china jadi pasar utama	ekspor.santan.surnel.tembus.543.ton.,.china.jadi.pasar.utama
1	wanita muda di Palembang ditemukan tewas di kamar hotel	wanita.muda.di.palembang.ditemukan.tewas.di.kamar.hotel
2	peringkatan akses keuangan daerah untuk perluat program prioritas pemerintah	peringkatan.akses.keuangan.daerah.untuk.perluat.program.prioritas.pemerintah
3	oknum brimob di opan lir dilaporkan liti ke program, ini dugaan kasusnya	oknum.brimob.di.opan.lir.dilaporkan.liti.ke.program,.ini.dugaan.kasusnya
4	kasus dugaan asusila oknum dosen terus bergulir, ini tanggapan bem km fisp unsi	kasus.dugaan.asusila.oknum.dosen.terus.bergulir,.ini.tanggapan.bem.km.fisp.unsi
5	pasikan pasokan listrik aman jelang porprov-peparprov surnel 2025 pin ulp sekayu cek sistem jaringan	pasikan.pasokan.listrik.aman.jelang.porprov-peparprov.surnel.2025.pin.ulp.sekayu.cek.sistem.jaringan
6	pokek bnt gelat patrol rufin antisipasi aksi balap liar	pokek.bnt.gelat.patrol.rufin.antisipasi.aksi.balap.liar
7	terciduk hendak mencuri, jurnaidi diamankan warga babatan saudagar	terciduk.hendak.mencuri,.jurnaidi.diamankan.warga.babatan.saudagar
8	kegiatan ilegal drilling di wilayah gunung megang dibongkar poles muara enim	kegiatan.ilegal.drilling.di.wilayah.gunung.megang.dibongkar.poles.muara.enim
9	kapolda surnel anggota polisi teribat narikoba bakal dipecat	kapolda.surnel.anggota.polisi.teribat.narikoba.bakal.dipecat
10	uji kemampuan dan kualitas, 15 mahasiswa prod pui un Rutu ujian munaqoyah	uji.kemampuan.dan.kualitas,.15.mahasiswa.prod.pui.un.Rutu.ujian.munaqoyah
11	motor digelapkan mantan suami, eva lapor polisi	motor.digelapkan.mantan.suami,.eva.lapor.polisi
12	agar proses pembelajaran berjalan sesuai kurikulum, guru man 1 disupervisi	agar.proses.pembelajaran.berjalan.sesuai.kurikulum,.guru.man.1.disupervisi
13	ppgpb tolak kenaikan tarif retribusi, kembalikan lapak puluhan pedang pasar jakabaring	ppgpb.tolak.keenaikan.tarif.retribusi,.kembalikan.lapak.puluhan.pedang.pasar.jakabaring
14	fm opnal macan puuh pokek bnt gelat patrol hunting sasar daerah rawan	fm.opnal.macan.puuh.pokek.bnt.gelat.patrol.hunting.sasar.daerah.rawan
15	karhutla hangusan 2 hektare lahan di jiwai oki	karhutla.hangusan.2.hektare.lahan.di.jiwai.oki
16	dua terdakwa kasus korupsi dpmd muba dituntut 3 dan 4 tahun penjara	dua.terdakwa.kasus.korupsi.dpmd.muba.dituntut.3.dan.4.tahun.penjara
17	dewa mnta hasil program pemberdayaan warga binaan lepas palembang di salurkan ke umkm	dewa.mnta.hasil.program.pemberdayaan.warga.binaan.lepas.palembang.di.salurkan.ke.umkm
18	pasikan jajan siswa sesuai standar kebersihan kepala madrasah sidak kanti sekolah	pasikan.jajan.siswa.sesuai.standar.kebersihan.kepala.madrasah.sidak.kanti.sekolah
19	sma 6 gandeng tni, bentuk karakter siswa dan cegah kenakalan remaja	sma.6.gandeng.tni,.bentuk.karakter.siswa.dan.cegah.kenakalan.remaja
20	keaji tetapkan dua tersangka kasus jaksa gadungan, salah satunya pns way kanan	keaji.tetapan.dua.terseangka.kasus.jaksa.gadungan,.salah.satunya.pns.way.kanan

Gambar. 1. Tokenizing

Index	tokenizing	stopword_removal
0	ekspor santan surnel tembus 543 ton, china jadi pasar utama	ekspor.santan.surnel.tembus.543.ton.,.china.pasar.utama
1	wanita muda di Palembang ditemukan tewas di kamar hotel	wanita.muda.palembang.ditemukan.tewas.kamar.hotel
2	peringkatan akses keuangan daerah untuk perluat program prioritas pemerintah	peringkatan.akses.keuangan.daerah.perluat.program.prioritas.pemerintah
3	oknum brimob di opan lir dilaporkan liti ke program, ini dugaan kasusnya	oknum.brimob.opan.lir.dilaporkan.liti.program,.dugaan.kasusnya
4	kasus dugaan asusila oknum dosen terus bergulir, ini tanggapan bem km fisp unsi	dugaan.asusila.oknum.dosen.terus.bergulir,.tanggapan.bem.km.fisp.unsi
5	pasikan pasokan listrik aman jelang porprov-peparprov surnel 2025 pin ulp sekayu cek sistem jaringan	pasikan.pasokan.listrik.aman.jelang.porprov-peparprov.surnel.2025.pin.ulp.sekayu.cek.sistem.jaringan
6	pokek bnt gelat patrol rufin antisipasi aksi balap liar	pokek.bnt.gelat.patrol.rufin.antisipasi.aksi.balap.liar
7	terciduk hendak mencuri, jurnaidi diamankan warga babatan saudagar	terciduk.hendak.mencuri,.jurnaidi.diamankan.warga.babatan.saudagar
8	kegiatan ilegal drilling di wilayah gunung megang dibongkar poles muara enim	kegiatan.ilegal.drilling.wilayah.gunung.megang.dibongkar.poles.muara.enim
9	kapolda surnel anggota polisi teribat narikoba bakal dipecat	kapolda.surnel.anggota.polisi.teribat.narikoba.dipecat
10	uji kemampuan dan kualitas, 15 mahasiswa prod pui un Rutu ujian munaqoyah	uji.kemampuan.kualitas,.15.mahasiswa.prod.pui.un.Rutu.ujian.munaqoyah
11	motor digelapkan mantan suami, eva lapor polisi	motor.digelapkan.mantan.suami,.eva.lapor.polisi
12	agar proses pembelajaran berjalan sesuai kurikulum, guru man 1 disupervisi	proses.pembelajaran.berjalan.sesuai.kurikulum,.guru.man.1.disupervisi
13	ppgpb tolak kenaikan tarif retribusi, kembalikan lapak puluhan pedang pasar jakabaring	ppgpb.tolak.keenaikan.tarif.retribusi,.kembalikan.lapak.puluhan.pedang.pasar.jakabaring
14	fm opnal macan puuh pokek bnt gelat patrol hunting sasar daerah rawan	fm.opnal.macan.puuh.pokek.bnt.gelat.patrol.hunting.sasar.daerah.rawan
15	karhutla hangusan 2 hektare lahan di jiwai oki	karhutla.hangusan.2.hektare.lahan.jiwai.oki
16	dua terdakwa kasus korupsi dpmd muba dituntut 3 dan 4 tahun penjara	terdakwa.korupsi.dpmd.muba.dituntut.3.4.penjara
17	dewa mnta hasil program pemberdayaan warga binaan lepas palembang di salurkan ke umkm	dewa.hasil.program.pemberdayaan.warga.binaan.lepas.palembang.salurkan.umkm
18	pasikan jajan siswa sesuai standar kebersihan kepala madrasah sidak kanti sekolah	pasikan.jajan.siswa.sesuai.standar.kebersihan.kepala.madrasah.sidak.kanti.sekolah
19	sma 6 gandeng tni, bentuk karakter siswa dan cegah kenakalan remaja	sma.6.gandeng.tni,.bentuk.karakter.siswa.dan.cegah.kenakalan.remaja
20	keaji tetapkan dua tersangka kasus jaksa gadungan, salah satunya pns way kanan	keaji.tetapan.terseangka.kasus.jaksa.gadungan,.salah.satunya.pns.way.kanan

Gambar. 2. Stopword Removal

Index	stopword_removal	stemming
0	ekspor santan surnel tembus 543 ton, china jadi pasar utama	ekspor.santan.surnel.tembus.543.ton.china.pasar.utama
1	wanita muda palembang ditemukan tewas kamar hotel	wanita.muda.palembang.temu.tewas.kamar.hotel
2	peringkatan akses keuangan daerah perluat program prioritas perintah	tingkat.akses.uang.daerah.perluat.program.prioritas.perintah
3	oknum brimob opan lir dilaporkan liti program, dugaan kasusnya	oknum.brimob.opan.lir.lapor.liti.program.duga.kasus
4	dugaan asusila oknum dosen gultit tanggap bem km fisp unsi	duga.asusila.oknum.dosen.gultit.tanggap.bem.km.fisp.unsi
5	pasikan pasokan listrik aman jelang porprov-peparprov surnel 2025 pin ulp sekayu cek sistem jaringan	pasikan.pasokan.listrik.aman.jelang.porprov-peparprov.surnel.2025.pin.ulp.sekayu.cek.sistem.jaring
6	pokek bnt gelat patrol rufin antisipasi aksi balap liar	pokek.bnt.gelat.patrol.rufin.antisipasi.aksi.balap.liar
7	terciduk hendak mencuri, jurnaidi diamankan warga babatan saudagar	ciduk.curi.jurnaidi.aman.warga.babatan.saudagar
8	kegiatan ilegal drilling wilayah gunung megang dibongkar poles muara enim	glat.ilegal.drilling.wilayah.gunung.megang.dibongkar.poles.muara.enim
9	kapolda surnel anggota polisi teribat narikoba bakal dipecat	kapolda.surnel.anggota.polisi.teribat.narikoba.pecat
10	uji kemampuan kualitas, 15 mahasiswa prod pui un Rutu ujian munaqoyah	uji.mampu.kualitas,.15.mahasiswa.prod.pui.un.Rutu.ujian.munaqoyah
11	motor digelapkan mantan suami, eva lapor polisi	motor.gelap.mantan.suami.eva.lapor.polisi
12	proses pembelajaran berjalan sesuai kurikulum, guru man 1 disupervisi	prosesajar.jalan.sesuai.kurikulum.guru.man.1.supervisi
13	ppgpb tolak kenaikan tarif retribusi, kembalikan lapak puluhan pedang pasar jakabaring	ppgpb.tolak.malk.tarif.retribusi.kembali.lapak.puluhan.pedang.pasar.jakabaring
14	fm opnal macan puuh pokek bnt gelat patrol hunting sasar daerah rawan	fm.opnal.macan.puuh.pokek.bnt.gelat.patrol.hunting.sasar.daerah.rawan
15	karhutla hangusan 2 hektare lahan jiwai oki	karhutla.hangusan.2.hektare.lahan.jiwai.oki
16	terdakwa korupsi dpmd muba dituntut 3 dan 4 tahun penjara	dakwa.korupsi.dpmd.muba.dituntut.3.4.penjara
17	dewa hasil program pemberdayaan warga binaan lepas palembang salurkan umkm	dewa.hasil.program.daya.warga.bina.lepas.palembang.salur.umkm
18	pasikan jajan siswa sesuai standar bersih kepala madrasah sidak kanti sekolah	pasikan.jajan.siswa.sesuai.standar.bersih.kepala.madrasah.sidak.kanti.sekolah
19	sma 6 gandeng tni, bentuk karakter siswa cegah kenakalan remaja	sma.6.gandeng.tni.bentuk.karakter.siswa.cegah.kenakalan.remaja
20	keaji tetapkan dua tersangka kasus jaksa gadungan, salah satunya pns way kanan	keaji.tetapan.sangka.jaksa.gadungan.salah.satunya.pns.way.kanan

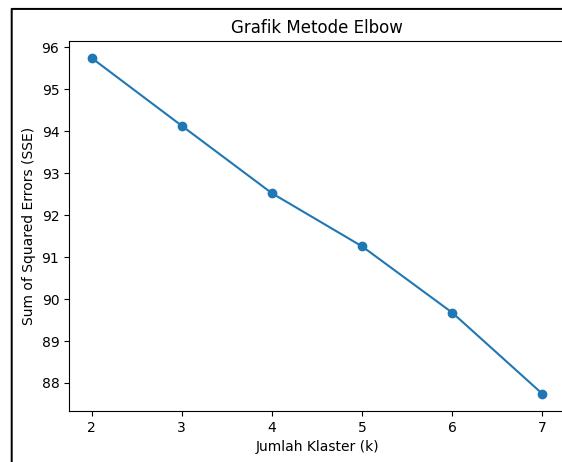
Gambar. 3. Stemming

Proses ini menghasilkan data teks yang lebih bersih dan terstandarisasi, sehingga siap digunakan pada tahap pembobotan kata menggunakan TF-IDF. Hasil pembobotan TF-IDF menunjukkan bahwa setiap judul berita berhasil direpresentasikan dalam bentuk vektor numerik yang mencerminkan tingkat kepentingan kata dalam dokumen, sebagaimana dijelaskan oleh Yusrian et al. (2020).

Index	11	148	15	17	2025	23	24	25	27	30	35	42	510	543	88	ajak	ajar	aksi	aman	anak
40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
41	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
42	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
43	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
44	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
45	0.0	0.0	0.0	0.0	0.30207447644128643	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
46	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3007219282288145	0.0	0.0	0.0	0.0
47	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
48	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
49	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3714563273157937	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Gambar. 4. TF-IDF

Penentuan jumlah kluster optimal dilakukan menggunakan Elbow Method. Hasil evaluasi menunjukkan adanya titik siku (elbow point) pada jumlah kluster tertentu, yang mengindikasikan jumlah kluster paling optimal untuk mengelompokkan judul berita.



Gambar. 5. Grafik Metode Elbow Penentuan Jumlah Kluster

Berdasarkan hasil tersebut, algoritma K-Means kemudian diterapkan untuk melakukan klusterisasi data judul berita.

cluster	count
0	47
1	14
6	10
4	9
3	8
5	6
2	6

dtype: int64

Gambar. 6. Distribusi Jumlah Judul Berita pada Setiap Kluster

Hasil klusterisasi menunjukkan bahwa judul berita suarapublik.id terbagi ke dalam beberapa kluster yang masing-masing merepresentasikan tema pemberitaan yang berbeda. Setiap kluster memiliki kata-kata dominan yang mencerminkan fokus topik tertentu, seperti isu pemerintahan, sosial, ekonomi, dan peristiwa daerah.

Temuan ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan judul berita berdasarkan kemiripan kata dan konteks tema secara efektif.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan pra-pemrosesan teks, pembobotan TF-IDF, dan algoritma K-Means mampu mengelompokkan judul berita pada website suarapublik.id ke dalam kluster yang merepresentasikan tema pemberitaan secara jelas dan sistematis. Hasil klusterisasi membuktikan bahwa data teks berupa judul berita dapat diubah menjadi representasi numerik yang efektif untuk mengidentifikasi pola dan kecenderungan topik pemberitaan media online lokal. Temuan ini memiliki implikasi teoretis dengan memperkuat landasan bahwa pendekatan text mining berbasis TF-IDF dan unsupervised learning dapat digunakan untuk menganalisis struktur tema pada data teks tidak terstruktur, khususnya dalam konteks media digital. Secara praktis, hasil penelitian ini dapat dimanfaatkan oleh pengelola media dan peneliti sebagai alat bantu analisis konten untuk memetakan fokus pemberitaan, mengevaluasi kecenderungan isu, serta mendukung pengambilan keputusan editorial berbasis data. Kontribusi ilmiah penelitian ini terletak pada penerapan

metode klusterisasi teks pada judul berita media online lokal, yang masih relatif terbatas dibandingkan penelitian sebelumnya yang umumnya berfokus pada dokumen akademik atau media nasional. Dengan demikian, penelitian ini melengkapi kekosongan kajian terkait analisis konten media lokal menggunakan pendekatan komputasional dan memperluas penerapan metode TF-IDF dan K-Means dalam kajian text mining. Penelitian lanjutan disarankan untuk memperluas objek analisis dengan menggabungkan judul dan isi berita secara keseluruhan, serta membandingkan performa algoritma K-Means dengan metode klusterisasi lain atau pendekatan berbasis topic modeling. Selain itu, eksplorasi penggunaan data temporal juga dapat dilakukan untuk menganalisis dinamika perubahan tema pemberitaan dari waktu ke waktu.

REFERENCES

- [1] Atimi RL, Pratama EE, Pradasari NI. Pendekatan Text Mining Untuk Klusterisasi Tren Penelitian Dengan Algoritma Unsupervised Learning K-Means. 2025;11(3):420–5.
- [2] Data S, Pada C, Detik S, Arta IM, Made G, Sasmita A, Et Al. Tf-Idf Dan K-Means Dengan Memanfaatkan. 2022;3(1).
- [3] Utami NW, Putra IGJE. Text Minig Clustering Untuk Pengelompokan Topik Dokumen Penelitian Menggunakan Algoritma K-Means Dengan Cosine Similarity. 2022;4(3):255–9.
- [4] Natalia Silalahi Glg. Bulletin Of Computer Science Research Rekomendasi Berita Berkaitan Dengan Menerapkan Algoritma Text Mining Dan TF-IDF. 2023;3(4):276–82.
- [5] Muthia Andini, Felicia Diana Rose, Joantika Lewis, Juan Reza Rizkilmy DR, Sinurat. Analisis Pengaruh AI : Perubahan Tingkat Kemalasan Mahasiswa Di Kota. 2024;10(2):95–103.
- [6] Nyoman Gede Yudiarta, Made Sudarma WGA. Pengelompokan Berita Pada Unstructured Textual Data. 2018;17(3):339–44.
- [7] Dwiki Krisnanda Wardya, I Ketut Gede Darma Putra NKDR. Clustering Artikel Pada Portal Berita Online. 2022;3(1):3–11.
- [8] Maria Ulfah ASI. Penerapan Data Mining Clustering Menggunakan Metode K-Means Dalam Pengelompokan Buku Perpustakaan Politeknik Negeri Balikpapan. 2022;4(3):62–8.
- [9] Mochammad Ilman Asnada, Bayu Rahayudi AR. Pengelompokan Topik Skripsi Mahasiswa Fakultas Ilmu Komputer Universitas Brawijaya Berdasarkan Judul Pada Periode 2015-2019 Menggunakan Metode Semi Supervised K-Means. 2022;6(1):58–65.
- [10] Rivaldo, R., Wibowo, A., & Mulyo H. Implementasi Algoritma K-Means Untuk Klusterisasi Data Hasil Tangkapan Ikan Di Karimunjawa. 2024;13:1045–56.
- [11] Syafina Haviyola, Susilawati MJ. Pengelompokan Prestasi Siswa Guna Kualifikasi Beasiswa Berdasarkan Data Nilai Menggunakan Algoritma K-Means. 2023;7(4):2786–91.
- [12] Jovansa Putra Laksana, Shela, Hafiz Irsyad Ar. Analisis Topik Dominan Dalam Paper Ilmu Komputer Menggunakan. 2025;3(3):78–84.
- [13] Eka Arriyanti, Siti Lailiyah Ak. Penerapan Metode K-Means Clustering Untuk Pengelompokan Judul Skripsi Mahasiswa (Studi Kasus : Stmik Widya Cipta Dharma). Apl Text Min Untuk Klusterisasi Aduan Masy Menggunakan Algoritm K-Means. 2021;
- [14] Dwi Remawati, Hendro Wijayanto, Yustina Retno Wahyu Utami Bdr. Pengelompokan Film Trending Di Youtube Menggunakan Tf-Idf Dan. 2025;4:65–74.
- [15] I Made Arta Purniawana, Gusti Made Arya Sasmita Ipaep. Tf-Idf Dan K-Means Dengan Memanfaatkan. 2022;3(1).
- [16] Adhitama, R., Prasetyo, A., & Nugraha D. Klusterisasi Judul Berita Pada Website Detik Menggunakan Algoritma Kmeans. 2024;1:194–207.