



# Minat Masyarakat Terhadap Karakteristik Informasi Tekstual dari Kementerian Lingkungan Hidup dan Kehutanan di Media Sosial

*Public interest of textual characteristics information of Ministry of Environment and Forestry on social media*

**Titis Hutama Syah<sup>a</sup>, Silvana Syah<sup>b</sup>, Nurmalasari<sup>c</sup>**

<sup>a</sup>Sekolah Tinggi Pertanian Kutai Timur, Jl. Soekarno-Hatta, Sangatta Utara, Kutai Timur, Kalimantan Timur, Indonesia

<sup>b</sup>Institut Bisnis dan Informatika Kosgoro 1957, Jl. Moch. Kahfi II No.33, Jagakarsa, Jakarta Selatan, Indonesia

<sup>c</sup>Universitas Nusa Mandiri, Jl. Jatiwaringin No. 2, Cipinang Melayu, Jakarta Timur, Indonesia

email: <sup>a</sup>titis@stiperkutim.ac.id, <sup>b</sup>deandraraisa@gmail.com, <sup>c</sup>nurmalasari.nmr@nusamandiri.ac.id

Menerima 5 November 2021, Revisi 12 Desember 2021, Diterima 5 Desember 2021, *Online* 7 Desember 2021

## ABSTRAK

Media sosial banyak dimanfaatkan masyarakat dalam memperoleh informasi, demikian juga lembaga pemerintahan, banyak yang menggunakan media sosial dalam menyebarkan informasi. Salah satunya adalah Kementerian Lingkungan Hidup dan Kehutanan (KLHK) melalui halaman Facebook *fanpage*-nya. Informasi yang diberikan pada umumnya berbentuk teks atau video dan gambar disertai penjelasan teks. Untuk itu, penelitian ini bertujuan untuk mengetahui karakteristik informasi tekstual terhadap ketertarikan minat pengunjung pada halaman Facebook *fanpage* KLHK. Penelitian dilakukan dengan cara melakukan *scraping* dan dilanjutkan pemrosesan bahasa alamiah (*Natural Language Processing*) untuk mengetahui karakteristik teks. Minat masyarakat diukur berdasarkan *likes*, *shares*, dan *comments*, dan disusun ke dalam klaster dengan metode Kmeans. Frekuensi kata yang sering muncul pada setiap klaster di analisis menggunakan similaritas *cosine* agar dapat dilakukan inferensi sesuai dengan domain KLHK. *Relabeling* dilakukan untuk menarik inferensi pada klaster yang terbentuk. Penelitian yang dilakukan dapat membentuk empat klaster. Klaster-klaster tersebut adalah klaster pengelolaan sumber daya alam nasional, pelestarian satwa liar, kegiatan pelestarian satwa liar, dan tim konservasi. Minat masyarakat tertinggi terdapat pada klaster pelestarian satwa liar yang ditunjukkan dengan jumlah *likes* terbanyak. Sedangkan klaster yang paling banyak dibagikan dan dikomentari adalah klaster tim konservasi. Informasi yang terdapat dalam penelitian ini dapat digunakan oleh lembaga terkait dalam meningkatkan dan mengembangkan penyebaran informasi melalui media sosial.

**Kata kunci :** Facebook *fanpage*, klasterisasi Kmeans, pemrosesan bahasa alamiah, pelestarian sumber daya alam. similaritas *cosine*

## ABSTRACT

Social media recently had widely used by public community in obtaining information. Likewise, many government institutions also use social media to disseminate information. One of them is the Ministry of Environment and Forestry (KLHK) through its Facebook *fanpage*. The information provided is mostly text or video and images accompanied by explanatory text. For this reason, this study aimed to determine the characteristics of textual information on visitor interest of the KLHK's Facebook fan page. The research was conducted by web *scraping* and continued with natural language processing (NLP) to find word characteristics. Public interest was measured based on *likes*, *shares*, and *comments*, and arranged into clusters using Kmeans method. Words frequencies that mostly appeared analysed using cosine similarity, therefore inference could be taken according to organisation's domain. *Relabeling* was done to draw general clusters inferences. The research conducted had been formed four clusters. The formed clusters were the national natural resource management, wildlife conservation, wildlife conservation activities, and the conservation team. The highest public interest was wildlife conservation cluster, which was indicated by the highest number of *likes*. Meanwhile, the cluster that was shared the most and commented on was conservation team cluster. The information contained in this research can be used by related institutions in improving and developing the dissemination of information through social media.

**Keyword :** cosine similarity, Facebook fan page, Kmeans clustering, natural language processing, natural resources conservation.

## PENDAHULUAN

Penggunaan media sosial dewasa ini telah berkembang cepat dan pemanfaatannya pun beragam, seperti: pengembangan citra individu, pengembangan dan perluasan jaringan komunitas, sarana perdagangan (jual-beli), pembentukan dan pembentukan citra produk dan organisasi, kampanye gagasan dan informasi, pertukaran opini, dan sebagainya. Aktifitas-aktifitas tersebut semakin berkembang seiring perkembangan teknologi informasi dan komunikasi. Kemudahan akses internet dan perangkat pendukung mengakibatkan peningkatan aktifitas di media sosial. Pengguna internet di Indonesia hingga Januari 2021 sebanyak 202,6 juta pengguna. Pengguna internet melalui perangkat bergerak/mobile internet (seperti: perangkat selular dan wifi) sebanyak 195,3 juta pengguna. Pengguna media sosial di Indonesia sebanyak 170 juta pengguna. Facebook merupakan salah satu media sosial yang banyak digunakan. Sebanyak 85,5% pengguna media sosial memanfaatkan Facebook dalam berkomunikasi (Kemp, 2021).

Media sosial juga dimanfaatkan oleh lembaga pemerintahan yang ada di Indonesia. Salah satunya adalah Kementerian Lingkungan Hidup dan Kehutanan (KLHK). KLHK memanfaatkan Facebook *fanpage* dalam menyebarkan informasi, menggunakan akun dengan nama Kementerian Lingkungan Hidup dan Kehutanan, yang dapat diakses melalui perambah dengan URL <https://facebook.com/KementerianLHK>, atau ditelusuri melalui aplikasi dengan nama akun tersebut. Akun tersebut resmi digunakan oleh KLHK, ditandai dengan centang biru setelah nama akun (🔵). Dalam dokumentasinya, Facebook menyatakan bahwa tanda tersebut merupakan tanda yang diberikan kepada akun yang valid dari seseorang atau perwakilan dari suatu organisasi (Facebook, 2021). Penggunaan media sosial di lembaga pemerintahan diatur dengan peraturan Menteri Pemberdayaan Aparatur Negara dan Reformasi Birokrasi (MenPAN-RB) Nomor 83 Tahun 2012. Media sosial oleh lembaga pemerintahan digunakan untuk menyebarkan informasi; membangun peran aparatur negara dan masyarakat; mensosialisasikan strategi dan tujuan pembangunan; membangun interaksi antara pemerintah dan masyarakat; meningkatkan kesadaran dan peran serta masyarakat; serta, menggali aspirasi, opini, dan masukan masyarakat.

Facebook memiliki komponen yang dapat digunakan untuk berbagi dan berinteraksi antara pemilik akun dan pengunjunnya. Hal-hal yang dibagikan dapat berupa teks, suara, gambar, dan video. Demikian juga akun KLHK. Halaman Facebook *fanpage* KLHK cenderung membagikan informasi melalui gambar atau video yang ditambahkan narasi atau deskripsi untuk menarik minat pengunjung. Pengunjung dapat menyukai (*like*), membagikan (*share*), atau mengomentari (*comment*) informasi tersebut. Jumlah dan bentuk respon yang muncul beragam sesuai karakteristik informasi yang diberikan. Permasalahan yang timbul adalah "Apakah setiap informasi menarik minat pengunjung?" dan "informasi seperti apa yang banyak diminati?". Eksplorasi terhadap pertanyaan-pertanyaan tersebut dapat membantu pemilik akun untuk mengembangkan penyebaran informasi sesuai fungsinya seperti yang tertera dalam peraturan MenPAN-RB Nomor 83 Tahun 2012.

Tulisan ini bertujuan untuk mengetahui karakteristik informasi tekstual terhadap ketertarikan minat pengunjung pada halaman Facebook *fanpage* KLHK. Fokus dari tulisan ini adalah ekstraksi informasi tertulis (*web scraping*) yang dikirimkan oleh pemilik akun, baik yang berdiri sendiri ataupun yang menjadi narasi dan deskripsi tentang gambar dan video. Luaran dari tulisan ini adalah model pengelompokan minat (diukur dengan *like*, *share*, dan *comment*) berdasarkan karakter-karakter teks yang disesuaikan dengan domain pengetahuan (*domain knowledge*) dan kebijakan dari KLHK sebagai lembaga pemerintahan. Informasi yang diperoleh dari penelitian ini diharapkan dapat memberikan masukan bagi KLHK sebagai lembaga pemerintahan dalam menggunakan media sosial dalam menyebarkan rencana dan capaian program kerja kepada masyarakat.

## KAJIAN PUSTAKA

Penyebaran (diseminasi) informasi merupakan kegiatan yang ditujukan kepada kelompok atau individu untuk memberikan informasi dengan maksud untuk menimbulkan kesadaran, penerimaan, dan pemanfaatan bagi penerimanya (Priliantini et al., 2018). Ragam cara penyebaran informasi dapat berupa penyampaian dari mulut ke mulut, penggunaan tanda dan simbol, media cetak, radio, televisi, internet, media online dan media sosial yang terus berkembang seiring dengan perkembangan peradaban dan teknologi (JP, 2021). Media sosial merupakan teknologi terkini yang banyak dimanfaatkan dalam upaya penyebaran informasi. Perkembangan media sosial di Indonesia sendiri tergolong cepat. Pengguna media sosial di Indonesia mencapai 170 juta jiwa (61,8 % dari jumlah total penduduk) pada Januari 2021, meningkat sebesar 6,3% dari periode sebelumnya (Kemp, 2021).

Media sosial merupakan sekelompok aplikasi berbasis Internet yang dibangun di atas dasar ideologi dan teknologi yang memungkinkan penggunaannya mempublikasikan informasi secara partisipatif dan kolaboratif atau biasa disebut web 2.0 (Kaplan & Haenlein, 2010). Media sosial juga memberikan ruang untuk berkreasi dan saling bertukar konten/informasi (*user generated content*). Berdasarkan definisi tersebut, aplikasi yang dapat dikategorikan sebagai media sosial antara lain adalah Wikipedia, YouTube, Facebook, Twitter, dan Instagram. Aplikasi-aplikasi tersebut dewasa ini dapat mudah ditemui dan digunakan pada gawai telepon pintar

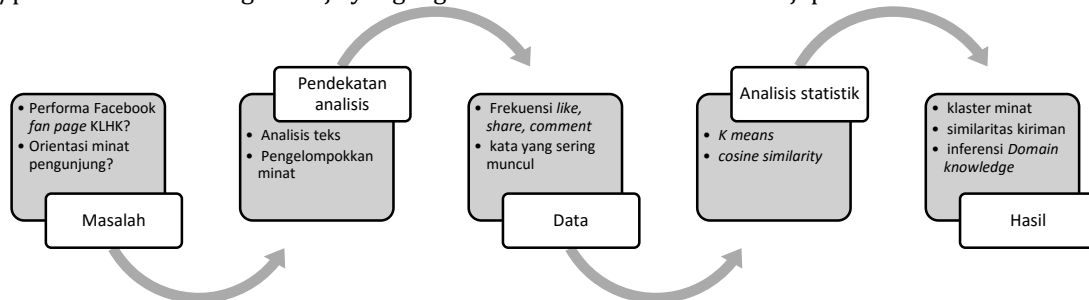
(*smartphone*). Aplikasi tersebut dapat digunakan oleh lembaga pemerintahan dalam berbagi (*sharing*) berita, program, pencapaian, atau untuk penyebarluasan informasi, serta membuka kesempatan bagi masyarakat untuk berpartisipasi dalam pembuatan keputusan dan mendengarkan opini masyarakat. Media sosial yang tepat untuk digunakan oleh lembaga pemerintahan adalah Facebook *fanpage* (Direktorat Jenderal Informasi dan Komunikasi Publik Kementerian Komunikasi dan Informatika, 2018).

Konten yang diterbitkan dalam media sosial agar dapat dianalisis diperlukan pra-pemrosesan menggunakan NLP (*Natural Language Processing*). Proses tersebut melibatkan aktifitas *normalization*, *POS tagging*, dan *parsing*. Dalam hal ini, *normalization* yang dimaksud adalah normalisasi teks (Louis, 2016). Dalam melakukan pra-pemrosesan teks, diperlukan aktifitas *tokenizing*, *stemming*, dan *stop word removal*. (Vijayarani & Janani, 2016). Informasi-informasi yang terdapat di media sosial dapat dianalisa secara statistik untuk memberikan rekomendasi artikel yang relevan bagi penggunanya. Hal ini dapat dilakukan dengan cara mengintegrasikan konten, informasi struktural, dan hubungan antar pengguna. Identifikasi diperlukan untuk membedakan karakteristik berbagai bentuk media sosial yang ada (Wang et al., 2010). Beberapa metode statistik dapat digunakan untuk menganalisis kinerja media sosial. Diantaranya adalah regresi, regresi logistik, analisis faktor, dan analisis klaster. Metode statistik ini biasanya memerlukan transformasi isi asli ke dalam format kode yang sesuai untuk metode statistik. Kerangka kerja untuk menganalisa konten yang terdapat di media sosial masih terbatas, dan memungkinkan pengembangan-pengembangan untuk metode baru (Zhang & Vos, 2014).

Tulisan ini memanfaatkan teknik pemrosesan teks yang dikombinasikan dengan pendekatan klusterisasi (*clustering*) untuk mengelompokkan minat pengunjung dari halaman penggemar (Facebook *fanpage*) KLHK untuk menentukan jenis kiriman yang diminati. Penelitian terdahulu tentang Facebook *fanpage* sudah banyak dilakukan. Antara lain adalah mengukur loyalitas pengunjung dengan menggunakan kuisisioner (Ruiz-Mafe et al., 2014), menggunakan kuisisioner untuk mengetahui niat (intensi) pengunjung dalam memberikan "like" pada suatu produk (Halaszovich & Nel, 2017), kategorisasi terhadap informasi-informasi yang diunggah dengan pendekatan regresi untuk menganalisisnya (Su et al., 2015).

## METODE PENELITIAN

Penelitian dilakukan dengan cara *web scraping* dari halaman Facebook *fan page* yang dapat diakses dari <http://m.facebook.com/KementerianLHK>. Ekstraksi informasi dilakukan menggunakan fasilitas facebook-scraapper versi 0.2.45 untuk bahasa pemrograman python (Zúñiga, 2021). Pengambilan informasi dilakukan pada tanggal 19 Agustus 2021 pada pukul 21.30 WIB. Informasi yang diperoleh meliputi kiriman teks (*post text*), *shared text*, waktu, gambar, ukuran gambar, deskripsi gambar, video, deskripsi video, durasi video, kualitas video, ukuran video, jumlah *likes-comments-shares*, informasi pengguna/pengunjung, teks *comments*, *reaction* dan jumlahnya. Informasi yang digunakan dalam penelitian ini adalah jumlah *likes-comments-shares* dan teks kiriman dari pengguna/pemilik akun. Kerangka kerja yang digunakan dalam tulisan ini tersaji pada Gambar 1.



Gambar 1. Kerangka kerja pengelompokan minat Facebook *fan page* KLHK

## Minat Pengguna/pengunjung

Minat pengguna/pengunjung diketahui berdasarkan banyaknya respon yang diperoleh dari suatu pesan/informasi yang dikirim, yaitu update status dari pemilik akun. Respon tersebut diukur banyaknya *like*, *comment*, dan *share* dengan asumsi bahwa pengunjung yang tertarik dengan informasi yang disampaikan oleh KLHK akan menyukainya, membagikannya melalui akun pribadi, atau mengomentarnya untuk berbagi opini. Respon yang terbentuk dikelompokkan dengan metode klusterisasi (*clustering*). Klusterisasi (*clustering*) adalah tindakan membagi data ke dalam kelompok-kelompok (klaster/cluster) dengan objek yang mirip. Metode ini dilakukan untuk menyederhanakan data dengan cara melakukan pengelompokan dengan pendekatan matematika, statistika, dan analisis numerik. Pencarian klaster merupakan pembelajaran tanpa pengawasan (*unsupervised learning*), dan hasilnya dapat mewakili konsep yang terkandung pada data (Berkhin, 2006).

Pendekatan klusterisasi dalam tulisan ini menggunakan metode K-means. K-means merupakan algoritma klusterisasi yang digunakan untuk pengumpulan data dalam jumlah besar (*big mining data*). K-means merupakan algoritma pembelajaran tanpa pengawasan (*unsupervised learning*), yang diterapkan untuk membentuk kluster. Metode ini digunakan untuk mengklasifikasikan objek ke dalam kluster yang berbeda secara berulang dan terpusat pada jarak data terdekat. Algoritma K-means terdiri dari dua langkah. Langkah pertama memilih pusat k secara acak yang nilainya ditetapkan terlebih dahulu. Langkah kedua adalah membawa setiap objek data ke setiap pusat yang terdekat. Jarak Euclidean digunakan untuk menentukan jarak antara setiap objek data dan pusat kluster. Proses pembentukan kluster dilakukan secara berulang hingga jarak objek data dengan kluster minimum (Shi et al., 2010). Persamaan-persamaan yang diperlukan dalam algoritma K-means (Shi et al., 2010) adalah sebagai berikut:

$$E = \sum_{i=1}^k \sum_{x \in C_i} |x - x_i|^2 \quad (1)$$

- $E$  : jumlah kuadrat error (*sum of square error / SSE*) seluruh objek
- $x$  : nilai objek data
- $x_i$  : rerata kluster  $C_i$
- $C_i$  : kluster ke - i (i= 1,2,3...)
- $k$  : banyaknya kluster

Jumlah kuadrat error menunjukkan variasi kesalahan dalam penentuan kluster. Sedangkan, jarak objek data dengan kluster ditentukan dengan jarak Euclidean yang digunakan untuk mencari jarak terdekat antara data dan pusat kluster (*centroid*). Persamaan jarak Euclidean adalah:

$$d(x_i, y_i) = \left[ \sum_{i=1}^n (x_i - y_i)^2 \right]^{1/2} \quad (2)$$

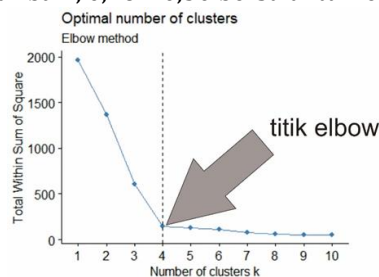
- $d$  : jarak euclidean
- $x_i, y_i$  : vector x dan y ke - i
- $n$  : banyaknya data

Pendekatan terhadap jumlah kluster yang dapat dibentuk digunakan metode *elbow*. Hal ini dilakukan untuk mempermudah penentuan jumlah kluster. Sedangkan efektifitas pembentukan kluster ditentukan berdasarkan koefisien *silhouette*. Pengulangan penentuan jumlah kluster terus dilakukan hingga tidak ada lagi data yang berpindah kluster. Untuk mempermudah hal tersebut dilakukan penyusunan grafik jumlah kluster berdasarkan nilai jumlah kuadrat error (SSE). Jika nilai sebuah kluster pertama dengan nilai kluster berikutnya membentuk sudut yang dalam pada grafik (membentuk pola seperti siku/*elbow*) atau nilainya mengalami penurunan paling besar maka jumlah nilai kluster tersebut dianggap tepat (Gambar 2). Cara tersebut merupakan prosedur yang terdapat dalam metode *elbow*. Sedangkan, koefisien *silhouette* digunakan untuk melihat kualitas dan kekuatan kluster, seberapa baik atau buruknya suatu obyek ditempatkan dalam suatu kluster (Dewi & Pramita, 2019). Persamaannya adalah:

$$SC = maks_k. SI(k) \quad (3)$$

- $SC$  : *Silhouette Coefficient* (koefisien *silhouette*)
- $SI$  : indeks *silhouette* global
- $k$  : jumlah kluster

Koefisien *silhouette* mendekati 1 berarti kluster yang terbentuk dapat memisahkan objek data dengan baik. Jika mendekati 0 berarti terdapat objek data yang bersinggungan. Sedangkan, nilai negatif menunjukkan terdapat percampuran objek data pada kluster yang terbentuk. Rentang dari nilai koefisien *silhouette* adalah 0,71 - 1,00 berstruktur kuat, 0,51 - 0,70 berstruktur baik, 0,26 - 0,50 berstruktur lemah, dan  $\leq 0,25$  berstruktur buruk.



Gambar 2. Penentuan jumlah kluster dengan metode *elbow* (Dewi & Pramita, 2019).

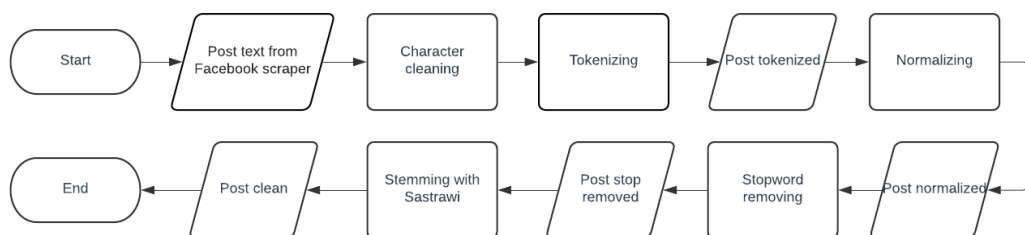
Prosedur-prosedur tersebut diatas diadopsi oleh modul komputasi *Scikit-learn* pada bahasa pemrograman python (Pedregosa et al., 2011). Pembentukan kluster terlebih dahulu melalui proses transformasi data dengan metode winsorisasi. Winsorisasi merupakan teknik transformasi yang mendekati *outlier* (pencilan) pada nilai persentil terdekat. Teknik ini tidak menghilangkan/menghapus data yang ada, mengingat informasi yang terkandung pada *outlier* masih diperlukan dalam analisis (Vandeput, 2021).

### Karakteristik informasi tekstual

Karakteristik informasi ditentukan berdasarkan kiriman teks (*post text*) yang diperoleh dari hasil *scraping*. Kiriman teks terlebih dahulu disederhanakan, yaitu menghilangkan karakter dan kata yang tidak diperlukan, link atau tautan, angka, tanda baca dan karakter tunggal. Setelah itu dilakukan tokenisasi (*tokenizing*), yaitu memisahkan struktur kalimat berdasarkan kata-kata penyusunnya (*token*). Kata-kata yang tidak baku dan salah tulis (*typo*) dinormalisasi menjadi bentuk baku. Penghilangan kata-kata tertentu (misalnya kata penghubung seperti: dan, atau, dari, dsb) yang dapat mempengaruhi pemrosesan teks (*stopword removal*). Kata-kata dengan frekuensi yang tinggi dan dianggap dapat mempengaruhi pemrosesan kata juga dihilangkan/diabaikan. *Stemming* merupakan proses terakhir dari penyederhanaan kiriman teks sebelum dihasilkan susunan teks baru yang terdiri dari rangkaian kata-kata dasar. *Stemming* sendiri merupakan tindakan menghilangkan imbuhan menjadi kata dasarnya saja. Proses tersebut dirangkum dalam Gambar 3. Algoritma yang digunakan adalah Sastrawi yang selain berperan untuk stemming juga mengubah seluruh huruf besar menjadi huruf kecil (*case folding*). Contoh dalam tahap ini adalah sebagai berikut:

kalimat: "Perekonomian Indonesia sedang dalam pertumbuhan yang membanggakan"  
disederhanakan menjadi: "ekonomi indonesia sedang dalam tumbuh yang bangga"

Alur kerja pemrosesan teks yang dilakukan terdapat pada Gambar 3.



Gambar 3. Alur kerja pemrosesan teks

Algoritma yang digunakan dalam pemrosesan teks merupakan bahasa pemrograman python. Modul NLTK (*Natural Language Toolkit*) digunakan untuk tokenisasi dan penghilangan *stopword* (Bird et al., 2009). Normalisasi kata disusun berdasarkan acuan *lexicon* (Salsabila et al., 2019). Proses *stemming* menggunakan modul Sastrawi versi 1.0.1 (Robbani, 2021).

*Post clean* (kata-kata yang sudah dibersihkan) dikelompokkan ke dalam kluster sesuai tingkat peminatannya. Namun hanya 20 kata dengan frekuensi paling banyak yang digunakan dalam analisis. Hubungan antar kata diukur tingkat similaritasnya untuk mengetahui kata-kata apa saja yang berperan besar dalam pembentukan kluster. Metode yang digunakan adalah *Cosine similarity*, biasanya digunakan untuk membandingkan dokumen atau memberikan peringkat dokumen berdasarkan vektor kata yang dihasilkan. Metode ini memanfaatkan fungsi cosinus (*Cosine*) untuk mengukur persamaan (similaritas) dua buah vektor (Han et al., 2012). Persamaan algoritma yang digunakan adalah:

$$sim(x, y) = \frac{x \cdot y}{\|x\| \|y\|} \quad (4)$$

Nilai  $\|x\|$  adalah jarak Euclidean dari vector  $x$ , dan nilai  $\|y\|$  adalah jarak Euclidean dari vector  $y$ . Nilai similaritas *Cosine* berkisar antara 0 hingga 1. Nilai 0 berarti bahwa kedua vector yang diukur tidak memiliki kesamaan. Sedangkan, nilai 1 bermakna vector yang diukur memiliki kesamaan.

Kata-kata yang memiliki similaritas tinggi digunakan untuk melakukan penamaan ulang kluster (*relabeling*) yang sudah terbentuk. Proses ini memerlukan pengetahuan tentang ruang gerak (*domain knowledge*) dari KLHK. Teknologi sistem informasi tidak hanya dapat berkembang dengan pengetahuan-pengetahuan tentang sistem itu sendiri, namun juga dapat dipadukan dengan pengetahuan dan teori spesifik tentang domain pengetahuan yang berbeda, seperti: humaniora, sains terapan, dan pembelajaran-pembelajaran interdisipliner lainnya (Hjørland &



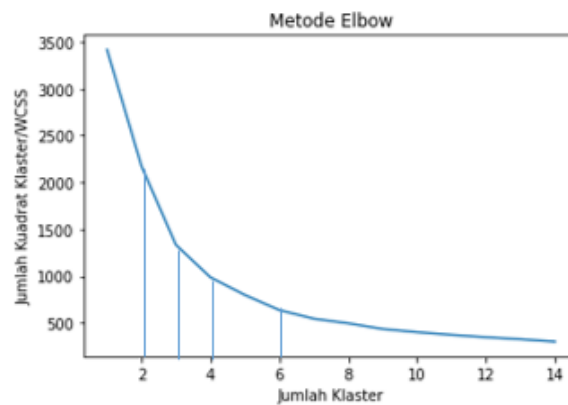
Albrechtsen, 1995). Sebagai bahan acuan dalam penentuan domain pengetahuan dari kluster yang terbentuk digunakan Rencana Strategis KLHK seperti yang tercantum pada peraturan Menteri LHK nomor P.16/MENLHK/SETJEN/SET.1/8/2020. Dengan demikian, klasterisasi yang dilakukan dapat dimanfaatkan oleh KLHK dalam mengembangkan lembaganya di media sosial.

## HASIL DAN PEMBAHASAN

Pengambilan data dengan secara *scraping* dari halaman Facebook *fanpage* KLHK diperoleh sebanyak 1139 informasi tekstual yang dikirimkan oleh pemilik akun. Kiriman paling awal yang dapat diambil tertanggal 4 juni 2020 pukul 11:29:29. Kiriman ini memperoleh 974 *likes*, dikomentari oleh 32 pengguna, dan dibagikan sebanyak 114 kali. Data terbaru yang berhasil diunduh adalah kiriman pada tanggal 19 agustus 2021 pukul 19:45:16. Hingga waktu pengunduhan, kiriman tersebut telah disukai sebanyak 22 pengguna, dikomentari oleh 1 pengguna, dan dibagikan sebanyak 3 kali. Kiriman yang paling disukai adalah kiriman pada tanggal 9 september 2020 pukul 19:27:09, sebanyak 16.404, dikomentari oleh 167 pengguna, namun tidak ada yang membagikan. Kiriman yang paling banyak dikomentari adalah kiriman pada tanggal 27 juni 2020 pukul 12:00:00. Kiriman yang paling banyak dibagikan adalah merupakan kiriman yang juga paling banyak dikomentari, dibagikan sebanyak 937 kali. Selengkapnya, disajikan pada Tabel 1.

Tabel 1.  
Variasi kiriman teks berdasarkan minat

Kiriman	Tanggal	Waktu	Jumlah likes	Jumlah comments	Jumlah shares	Kata-kata awal
Paling awal	4 juni 2020	11:29:29	974	32	114	Eggang gading adalah jenis burung rangkong...
Paling akhir	19 agustus 2021	19:45:16	22	1	3	KLHK mengapresiasi kiprah bank sampah dalam ...
Paling banyak likes	9 september 2020	19:27:09	16404	167	0	Hai dapat salam nih dari Parama si anak elang ...
Paling banyak comments	27 juni 2020	12:00:00	879	360	937	Masih banyak nih yang bertanya bagaimana cara ...
Paling banyak shares	27 juni 2020	12:00:00	879	360	937	Masih banyak nih yang bertanya bagaimana cara ...

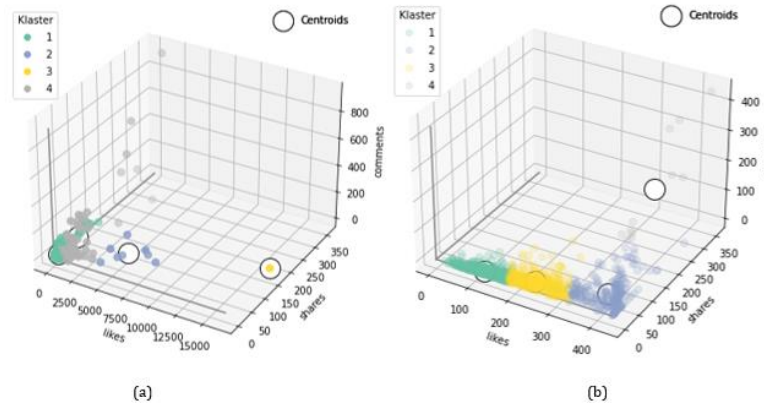


Gambar 4. Kurva elbow untuk menentukan kluster

Tabel 2.  
Koefisien silhouette jumlah kluster

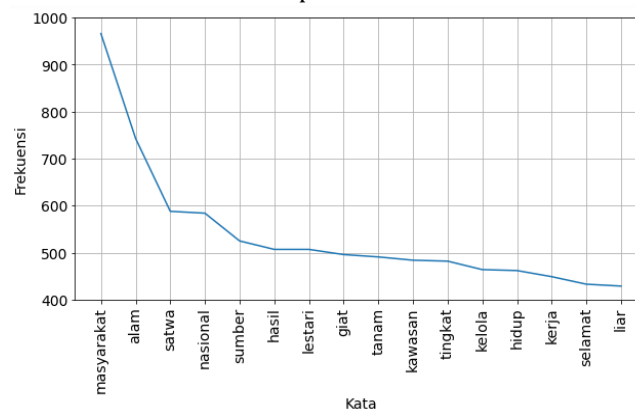
Jumlah kluster	Koefisien silhouette	Struktur
2	0.91	Kuat
3	0.90	Kuat
<b>4</b>	<b>0.71</b>	<b>Kuat</b>
5	0.66	Baik
6	0.55	Baik

Minat atau ketertarikan pengunjung diindikasikan dengan banyaknya like, comment dan share. Asumsi yang diangkat dalam tulisan ini adalah ketiga indikator tersebut memiliki kemiripan yang dapat membentuk kluster. Pendekatan menggunakan metode elbow menunjukkan bahwa kluster yang dapat terbentuk dan dapat digunakan untuk klusterisasi dengan kmeans adalah sebanyak 4 ( $k = 4$ ). Gambar 4 memperlihatkan bahwa ada beberapa sudut tajam yang terbentuk dalam kurva elbow. Untuk mengetahui akurasi dari jumlah kluster dilakukan pendekatan dengan menghitung koefisien silhouette. Tabel 2 menunjukkan bahwa kluster yang mampu memberikan struktur yang kuat adalah sebanyak 4 kluster atau kurang, yaitu nilai koefisien yang lebih besar dari atau sama dengan 0,71 (Dewi & Pramita, 2019). Oleh karena itu, jumlah kluster yang digunakan untuk klusterisasi dengan kmeans adalah sebanyak 4 kluster.



Gambar 5. Sebaran data dan centroid minat pengunjung Facebook fan page KLHK. (a) sebelum ditransformasi, (b) setelah ditransformasi,

Empat kluster yang terbentuk terdiri atas 449 kiriman yang termasuk kluster 1, sebanyak 278 kiriman pada kluster 2, sebanyak 401 pada kluster 3, dan 11 kiriman pada kluster 4. Pembentukan kluster ini terlebih dahulu melalui proses transformasi data dengan metode winsorisasi (Vandeput, 2021). Persentil yang digunakan untuk mentransformasi outlier adalah persentil ke 95 atau sebanyak 5% data dengan nilai tertinggi dianggap sebagai outlier. Meskipun demikian, transformasi outlier dilakukan dengan tujuan untuk membentuk kluster yang kuat. Gambar 5 menunjukkan perbedaan sebaran data dan centroid sebelum dan setelah ditransformasi. Gambar tersebut menunjukkan bahwa sebaran data yang telah ditransformasi dapat membentuk kalster lebih baik, ditandai dengan tumpang tindih data antar kluster dapat diminimalkan.



Gambar 6. Dua puluh kata yang sering muncul pada kiriman Facebook fan page KLHK.

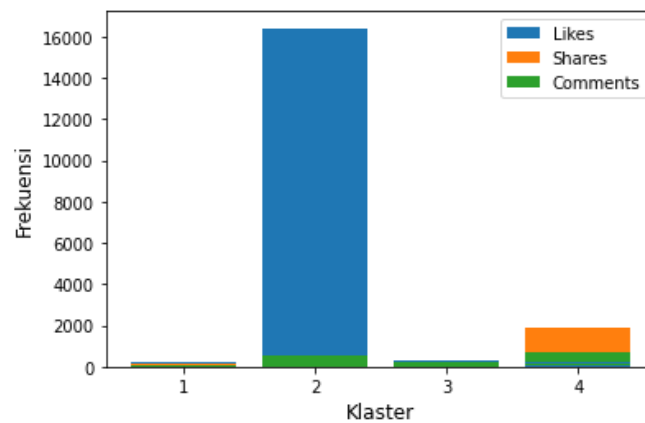
Kluster yang terbentuk selanjutnya dilakukan relabeling berdasarkan karakteristik informasi. Pendekatan yang dilakukan adalah menggunakan cosine similarity. Similaritas yang diukur adalah 20 yang paling banyak muncul pada setiap kluster. Secara umum, 20 kata yang paling banyak muncul dari seluruh kiriman teks tersaji pada Gambar 6. Sedangkan, similaritas kata yang paling banyak muncul pada setiap kluster tersaji pada Gambar 7. Dari keterikatan kata yang terdapat dalam kluster dilakukan relabeling informasi sesuai struktur hubungan kata yang terbentuk.

Tabel 3 menunjukkan kata-kata yang sering muncul pada setiap kluster. Untuk kluster 1, hubungan antara kedua kata yang sering muncul adalah "masyarakat : kelola" dan "alam : nasional". Berdasarkan nilai similaritasnya kombinasi kata-kata tersebut muncul pada 40% kiriman yang termasuk dalam kluster 1. Pada kluster 2, kombinasi

kata yang sering muncul adalah “satwa : liar” dan “alam : lestari” (60 % kiriman). Pada klaster 3 kombinasi yang sering muncul adalah “alam : lestari” dan “giat : laksana” (60% kiriman). Sedangkan, pada klaster 4, kombinasi “satwa : tim” dan “konservasi : tim” sering muncul pada 85 % kiriman yang termasuk anggota klaster. Berdasarkan kombinasi-kombinasi tersebut, relabeling dilakukan dan disesuaikan dengan fungsi dan domain tugas KLHK sebagai lembaga pemerintahan. Distribusi minat pengunjung tersaji pada Gambar 8.

Tabel 3.  
Kombinasi kata dengan similaritas tertinggi

Klaster	Jumlah Kiriman	Kombinasi kata yang sering muncul	Cosine similarity	Relabeling
1	449	masyarakat : kelola alam : nasional (Similaritas kemunculan 40%)	0.520886 0.425090	Pengelolaan sumberdaya alam nasional
2	278	satwa : liar alam : lestari (Similaritas kemunculan 60%)	0.674330 0.657199	Pelestarian satwa liar
3	401	alam : lestari giat : laksana (Similaritas kemunculan 60%)	0.610483 0.631364	Pelaksanaan kegiatan pelestarian alam
4	11	satwa : tim konservasi : tim (Similaritas kemunculan 85%)	0.866025 0.866025	Tim konservasi

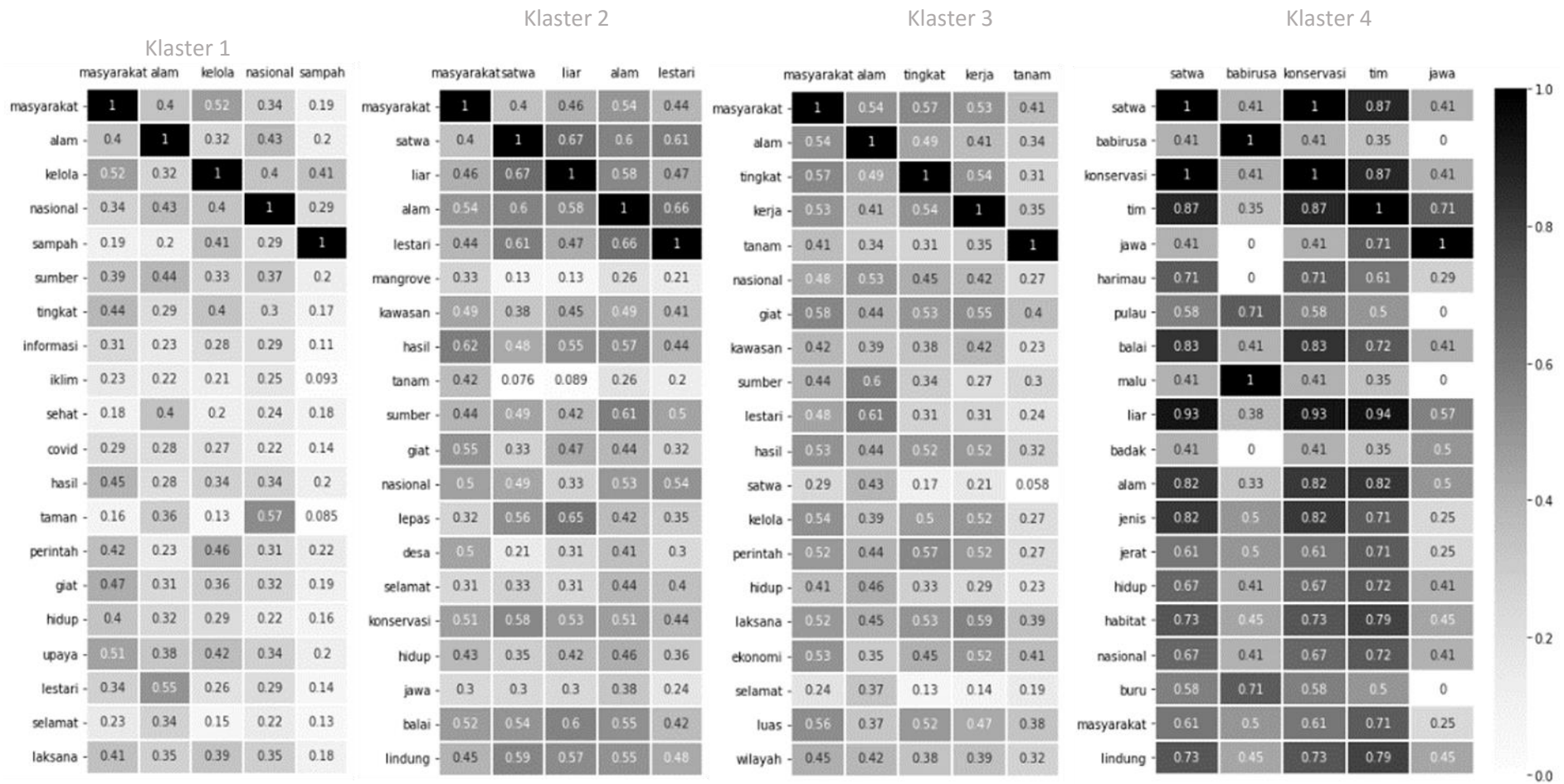


Gambar 8. Distribusi minat pengunjung

### Minat pengunjung Facebook fanpage KLHK

Karakter teks yang disukai, dikomentari, dan dibagikan oleh pengunjung berusaha dipetakan, namun tidak sepenuhnya dapat dikelompokkan dalam klaster yang terbentuk. Tabel 3 menunjukkan bahwa kiriman yang menarik minat pengunjung dapat terbagi dalam 4 klaster, meskipun demikian tidak semua kiriman dapat dijadikan inferensi yang utuh. Klaster 1 menunjukkan bahwa hanya sebanyak 40 % kiriman yang dapat ditarik satu kesimpulan yang sama. Sisanya sebesar 60% menunjukkan kehadiran teks yang beragam dan kemiripan teks antar kiriman yang rendah, atau merupakan kejadian acak untuk dapat dijadikan inferensi bahwa pengunjung tertarik pada kiriman tersebut. Hal ini juga terjadi pada klaster – klaster lainnya dengan persentase similaritas yang berbeda. Kejadian ini menurut merupakan kelemahan dari analisis isi (*content analysis*) yang lebih menekankan pada frekuensi kemunculan teks dan tidak dapat sepenuhnya menangkap makna yang terkandung (Rahardjo, 2010). Pendekatan terhadap makna yang terkandung dalam kiriman teks dilakukan dengan melakukan *relabeling*, yaitu: klaster 1 dikelompokkan sebagai informasi tentang ‘Pengelolaan sumberdaya alam nasional’, klaster 2 tentang ‘Pelestarian satwa liar’, klaster 3 tentang ‘Pelaksanaan kegiatan pelestarian alam’, dan klaster 4 tentang ‘Tim konservasi’. Setidaknya, tulisan ini dapat memberikan informasi tentang kecenderungan minat pengunjung terhadap informasi yang diberikan terhadap klaster yang terbentuk. Gambar 8 menunjukkan bahwa informasi yang banyak menarik perhatian pengunjung adalah tentang satwa liar, sedangkan yang kurang menarik adalah terkait dengan pengelolaan sumber daya alam.





Gambar 7. Matriks (20 x 5) cosine similarity setiap kluster berdasarkan kata yang sering muncul

KLHK sesuai dengan rencana kerjanya memiliki 4 (empat) sasaran strategis yang menjadi isu utama, yaitu: perubahan iklim, pemanfaatan sumber daya hutan dan lingkungan, distribusi manfaat hutan dan tata Kelola dan Inovasi. Terdapat indikator-indikator yang digunakan untuk mengukur keberhasilannya. Penyebaran informasi juga merupakan salah satu kegiatan yang disebutkan pada rencana kegiatan KLHK. Indikator yang digunakan adalah jumlah kunjungan ke *website* resmi KLHK. Dengan demikian, penggunaan media sosial belum menjadi prioritas utama dalam penyebaran informasi. Hal ini dapat menjadi sebab kiriman teks pada Facebook *fan page* KLHK masih sebatas digunakan untuk menarik minat pengguna Facebook saja, namun belum digunakan sebagai media kampanye rencana dan capaian kerja lembaganya secara menyeluruh. KLHK banyak menggunakan isu tentang satwa liar untuk menarik minat pengguna Facebook. Meskipun demikian, isu satwa liar sendiri merupakan salah satu komponen yang terdapat dalam indikator perubahan iklim yang menjadi sasaran KLHK. *Social networking sites* (SNS) dapat digunakan pemerintah untuk memberikan dan meningkatkan transparansi, kolaborasi, dan partisipasi masyarakat dalam mengambil kebijakan. Keberadaan SNS, seperti Facebook, dapat digunakan untuk menggali opini, umpan balik, dan ide-ide dari masyarakat. Pemanfaatan media sosial secara lebih komprehensif dapat digunakan oleh lembaga pemerintahan untuk mempercepat pelaksanaan pembangunan (Wirtz et al., 2017). Fungsi ini dapat dikembangkan oleh KLHK dalam akun Facebook *fan page*-nya untuk lebih dapat menyerap aspirasi masyarakat terhadap kinerja kelembagaan dan meningkatkan fungsi media sosial sebagai sarana utama dalam mempermudah penyebaran informasi.

## KESIMPULAN

Media sosial yang digunakan oleh KLHK salah satunya adalah Facebook *fan page*. Tipe informasi yang berbentuk kiriman teks dapat dikelompokkan menjadi 4 klaster. Inferensi terhadap domain pemilik akun menunjukkan ke empat klaster tersebut dapat dikelompokkan menjadi klaster pengelolaan sumber daya alam nasional, pelestarian satwa liar, pelaksanaan kegiatan pelestarian alam, dan informasi tentang tim konservasi. Informasi yang banyak diminati oleh pengunjung adalah tentang satwa liar, meskipun KLHK memiliki ruang gerak yang lebih banyak berdasarkan rencana dan strategi kerjanya. Didalam rencana kerjanya, KLHK belum memprioritaskan penggunaan media sosial dalam penyebaran informasi, seperti yang ditunjukkan distribusi klaster yang terbentuk.

Pemetaan minat karakteristik informasi Facebook *fan page* KLHK dilakukan dengan pendekatan klasterisasi Kmeans dan similaritas *cosine*. Terdapat bias yang muncul terutama pada proses inferensi domain pengguna berdasarkan frekuensi kemunculan kata. Inferensi memberikan bias karena tidak dapat mengakomodasikan distribusi kiriman teks secara utuh. Metode yang digunakan masih memerlukan pengembangan untuk meminimalisir bias. Pendekatan dengan metode lain dapat dilakukan, diantaranya adalah *angular separation*, *manhattan distance*, dan *euclidean distance* untuk mengukur similaritas karakter tekstual yang terbentuk.

## Ucapan Terima Kasih

Terima kasih penulis sampaikan kepada Kementerian Komunikasi dan Informasi yang telah menyelenggarakan pelatihan thematic academy data scientist, sehingga penulis dapat menyusun tulisan ini.

## DAFTAR PUSTAKA

- Berkhin, P. (2006). A Survey of Clustering Data Mining Techniques. *Grouping Multidimensional Data: Recent Advances in Clustering*, 25–71.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python* (J. Steele, Ed.). O'Reilly Media, Inc. <http://www.datascienceassn.org/sites/default/files/Natural%20Language%20Processing%20with%20Python.pdf>
- Dewi, D. A. I. C., & Pramita, D. A. K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix : Jurnal Manajemen Teknologi Dan Informatika*, 9(3), 102–109.
- Direktorat Jenderal Informasi dan Komunikasi Publik Kementerian Komunikasi dan Informatika. (2018). *Memaksimalkan Penggunaan Media Sosial Dalam Lembaga Pemerintah*. Direktorat Jenderal Informasi dan Komunikasi Publik, Kementerian Komunikasi dan Informatika. [https://indonesiabaik.id/public/uploads/post/2997/Memaksimalkan\\_Penggunaan\\_Media\\_Sosial\\_dalam\\_Lembaga\\_Pemerintah.pdf](https://indonesiabaik.id/public/uploads/post/2997/Memaksimalkan_Penggunaan_Media_Sosial_dalam_Lembaga_Pemerintah.pdf)
- Halaszovich, T., & Nel, J. (2017). Customer–brand engagement and Facebook fan-page “Like”-intention. *Journal of Product & Brand Management*, 26(2), 120–134.

- Han, J., Kamber, M., & Pei, J. (2012). Getting to Know Your Data. *Data Mining*, 39–82. <https://doi.org/10.1016/B978-0-12-381479-1.00002-2>
- Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain-analysis. *Journal Of The American Society For Information Science*, 46(6), 400–425.
- JP, S. (2021, June 28). *Perkembangan Platform Media Penyebaran Informasi*. <https://kompaspedia.kompas.id/baca/infografik/poster/perkembangan-platform-media-penyebaran-informasi>
- Kaplan, A., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53, 59–68.
- Kemp, S. (2021, February 11). *Digital in Indonesia: All the Statistics You Need in 2021 — DataReportal – Global Digital Insights*. <https://datareportal.com/reports/digital-2021-indonesia>
- Louis, A. (2016). Natural Language Processing for Social Media. *Computational Linguistics*, 42(4), 833–836.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830.
- Rahardjo, M. (2010, November 15). *Mengenal Ragam Studi Teks: Dari Content Analysis hingga Pos-modernisme*. <https://www.uin-malang.ac.id/r/101101/mengenal-ragam-studi-teks-dari-content-analysis-hingga-pos-modernisme.html>
- Robbani, H. A. (2021). *Sastrawi · PyPI* (1.0.1). <https://pypi.org/project/Sastrawi/>
- Ruiz-Mafe, C., Martí-Parreño, J., & Sanz-Blas, S. (2014). Key drivers of consumer loyalty to Facebook fan pages. *Online Information Review*, 38(3), 362–380.
- Salsabila, N. A., Ardhito Winatmoko, Y., Akbar Septiandri, A., & Jamal, A. (2019). Colloquial Indonesian Lexicon. *Proceedings of the 2018 International Conference on Asian Language Processing, IALP 2018*, 226–229.
- Shi, N., Liu, X., & Guan, Y. (2010). Research on k-means clustering algorithm: An improved k-means clustering algorithm. *3rd International Symposium on Intelligent Information Technology and Security Informatics, IITSI 2010*, 63–67.
- Su, N., Reynolds, D., & Sun, B. (2015). How to make your Facebook posts attractive: A case study of a leading budget hotel brand fan page. *International Journal of Contemporary Hospitality Management*, 27(8), 1772–1790.
- Vandeput, N. (2021). Data Science for Supply Chain Forecasting. *Data Science for Supply Chain Forecasting*. <https://doi.org/10.1515/9783110671124/HTML>
- Vijayarani, S., & Janani, R. (2016). Text Mining: open Source Tokenization Tools – An Analysis. *Advanced Computational Intelligence: An International Journal (ACIJ)*, 3(1), 37–47.
- Wang, J., Li, Q., Chen, Y. P., & Lin, Z. (2010). Recommendation in Internet Forums and Blogs. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 257–265. <https://aclanthology.org/P10-1027>
- Wirtz, B. W., Daiser, P., & Mermann, M. (2017). Social Media as a Leverage Strategy for Open Government: An Exploratory Study. <https://doi.org/10.1080/01900692.2017.1289388>, 41(8), 590–603.
- Zhang, B., & Vos, M. (2014). Social media monitoring: aims, methods, and challenges for international companies. *Corporate Communications: An International Journal*, 19(4), 371–383.