

Analisis Sentimen Ulasan Wisatawan Terhadap Destinasi Gili di KLU Menggunakan *K-Means* Clustering

(*Sentiment Analysis of Tourist Reviews of Gili Destinations in KLU Using K-Means Clustering*)

Juanri Priskila Obenu^[1], I Gede Pasek Suta Wijaya^[2], Fitri Bimantoro^[3]

^[1]Dept Informatics Engineering, Mataram University

Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA

Email: juanripriskilaaa@gmail.com, [gpsutawijaya, bimo]@unram.ac.id

Abstract

The rise of social media, particularly Twitter, has enabled sentiment analysis in tourism. This study examines tourist sentiments toward Gili, Lombok, using K-Means Clustering on 6,030 tweets, refined to 3,156 after cleaning. The clustering yielded Silhouette Coefficient scores of 0.4382 (positive-English), 0.4173 (positive-Indonesian), 0.4258 (negative-English), and 0.4535 (negative-Indonesian), indicating well-structured clusters. Tourists expressed positive views on natural beauty and activities but raised concerns about water supply issues. This study demonstrates K-Means Clustering's effectiveness in revealing sentiment trends, offering valuable insights for local governments and tourism stakeholders to enhance visitor experiences and address challenges.

Keywords: Sentiment Analysis, Text Mining, Clustering, K-Means

1. PENDAHULUAN

Pulau Lombok merupakan bagian dari wilayah administratif Provinsi Nusa Tenggara Barat (NTB), sehingga menjadi pilihan utama untuk pengembangan sektor pariwisata di NTB dibandingkan dengan pulau-pulau lain dalam wilayah tersebut. tahun 2015, Lombok meraih penghargaan World Travel Halal Summit di Abu Dhabi, Uni Emirat Arab [1]. Prestasi ini secara tidak langsung mendorong peningkatan kunjungan wisatawan ke NTB, terutama ke Pulau Lombok. Selain itu, semakin banyak destinasi wisata yang kini lebih mudah diakses, baik dari segi informasi, transportasi, maupun akomodasi. Faktor ini menjadi pertimbangan penting bagi wisatawan yang ingin berkunjung ke Lombok [2].

Berdasarkan Peraturan Daerah (PERDA) Provinsi NTB Nomor 7 Tahun 2013, telah ditetapkan Kawasan Strategis Pariwisata Daerah (KSPD) yang mencakup Destinasi Pariwisata Daerah (DPD) Lombok. Di Pulau Lombok sendiri, terdapat empat KSPD utama, yaitu kawasan Mataram Metro, kawasan Senggigi – Tiga Gili, kawasan Kuta - Mandalika, dan kawasan Rasimas – Sembalun. Saat ini, kawasan wisata Senggigi – Tiga Gili masih menjadi favorit wisatawan karena letaknya yang relatif dekat dari pusat kota serta akses transportasi yang beragam dan memadai. Kawasan Tiga Gili, yang terletak di Kabupaten Lombok Utara, terdiri dari Gili

Trawangan, Gili Meno, dan Gili Air. Gili Trawangan dan Gili Air merupakan kawasan yang sudah cukup padat penduduk dan pengunjung, menjadikannya destinasi unggulan bagi wisatawan yang mencari pengalaman wisata bahari yang menarik [2].

Pertumbuhan media sosial, terutama Twitter, sejalan dengan kemajuan teknologi informasi, menjadikannya sebagai platform informasi yang populer untuk sosialisasi online. Masyarakat memanfaatkan Twitter untuk menyampaikan opini, kritik, dan saran secara bebas, serta mendapatkan informasi secara cepat dan *real-time* [3]. Twitter berfungsi sebagai sumber informasi penting untuk menggali sentimen masyarakat melalui cuitan mengenai berita atau peristiwa tertentu terkait pariwisata Gili di KLU. Meskipun informasi yang ditransmisikan melalui Twitter sangat berharga, analisis manual menjadi sulit karena volume *tweet* yang sangat besar. Oleh karena itu, diperlukan model analisis data yang efisien, seperti analisis sentimen dan pemodelan topik, untuk menemukan informasi tersembunyi dalam kumpulan *tweet* [3].

Analisis sentimen merupakan bagian dari *text mining* yang bertujuan untuk mengklasifikasikan dokumen teks berupa opini menjadi makna positif atau negatif yang berguna untuk menganalisis pendapat dan emosi terhadap topik tertentu [4]. Hasil analisis ini memberikan wawasan mengenai kepuasan wisatawan,

yang dapat digunakan oleh pemerintah daerah dan pelaku industri pariwisata untuk merancang kebijakan yang lebih tepat sasaran [5]. Namun, untuk mengelola volume data yang besar dan kompleksitas bahasa, diperlukan proses pengelompokan atau *clustering* agar data menjadi lebih terstruktur dan mudah dipahami.

Clustering adalah teknik untuk mengelompokkan data berdasarkan kesamaan atau kecocokan, yang bertujuan untuk memahami pola dan karakteristik dalam dataset yang besar. Dengan menggunakan *clustering*, data yang belum terolah dapat dibagi menjadi beberapa grup yang lebih mudah dianalisis [6]. *K-Means* merupakan metode *clustering* yang sering digunakan karena kemampuannya dalam mengelompokkan data secara efisien berdasarkan kedekatannya dengan *centroid* [7]. Sehingga dalam konteks ini, *K-Means* digunakan untuk mengelompokkan opini wisatawan berdasarkan intensitas sentimen mereka, seperti sentimen positif dan negatif, serta membagi setiap kategori menjadi *cluster* yang lebih spesifik. Hal ini mempermudah dalam mengidentifikasi pola sentimen dan memahami nuansa emosi yang terkandung dalam ulasan pariwisata di Gili, KLU, sehingga dapat memberikan wawasan yang lebih mendalam mengenai tingkat kepuasan atau ketidakpuasan wisatawan.

2. TINJAUAN PUSTAKA DAN TEORI PENUNJANG

2.1 Tinjauan Pustaka

Berikut merupakan hasil tinjauan literatur yang telah dilakukan pada sejumlah jurnal penelitian sebelumnya terkait dengan topik metode *clustering K-Means*, yang dapat dijelaskan sebagai berikut:

Penelitian [8] menunjukkan bahwa algoritma *Support Vector Machine* (SVM) bekerja dengan baik dalam analisis sentimen, dengan rata-rata akurasi sebesar 74,39%. Selain itu, hasil *clustering* dengan *K-Means* sangat dipengaruhi oleh nilai titik pusat *cluster* (*centroid*) dan jumlah data yang digunakan. Perbedaan dalam pemilihan titik pusat *cluster* awal juga berkontribusi terhadap variasi hasil pengelompokan.

Penelitian [9] menunjukkan bahwa analisis sentimen berbasis klaster dapat membantu mengelompokkan skor sentimen berdasarkan polaritas dan subjektivitas. *Pre-defined dictionaries* atau alat analisis sentimen sering kali tidak dapat menangkap skor yang tepat untuk setiap kata dalam konteks kalimat. Oleh karena itu, dengan membentuk klaster berdasarkan skor dari berbagai alat, penelitian ini berhasil mengidentifikasi tweet yang benar-benar positif dan negatif. Selain itu, dengan menganalisis subjektivitas—yakni opini, perasaan, atau emosi dalam

suatu tweet—klaster dapat dibentuk untuk membedakan antara opini nyata dan sekadar ekspresi sentimen. Studi ini diterapkan pada data spesifik terkait IPL 2017, dengan rencana untuk menyesuaikan label dan skor sentimen menggunakan *classifier* yang lebih kontekstual.

Penelitian [10] menekankan pentingnya preprocessing data dalam analisis sentimen, di mana *K-Means Clustering* digunakan untuk meningkatkan struktur data, dan SMOTE digunakan untuk menangani ketidakseimbangan kelas. Hasil penelitian menunjukkan bahwa metode dasar dengan *Random Forest* memiliki akurasi 77,49%, namun setelah diterapkan *K-Means*, akurasinya meningkat sebesar 20,98%. Peningkatan tertinggi terjadi pada kombinasi SMOTE-*K-Means*, yang meningkatkan akurasi sebesar 22,06% serta meningkatkan *precision*, *recall*, dan F1-score secara signifikan. Hasil ini membuktikan bahwa integrasi *K-Means* dan SMOTE mampu meningkatkan performa klasifikasi sentimen secara efektif.

Penelitian [11] menunjukkan bahwa *K-Means++* menghasilkan klaster yang lebih baik dibandingkan *K-Means*, terutama untuk nilai *k* yang lebih tinggi, dengan *Silhouette Coefficient* (SC) rata-rata 0,882 pada *k* = 5. Sementara itu, *K-Means* lebih efektif untuk *k* yang lebih rendah, dengan SC 0,825 pada *k* = 3 dan 0,873 pada *k* = 4. Hasil klasterisasi menggunakan *K-Means++* mengidentifikasi Jakarta Barat, Jakarta Selatan, Jakarta Utara, dan Jakarta Pusat sebagai wilayah sangat rawan penyebaran Covid-19, sedangkan Kab. Bogor, Surabaya, dan Jakarta Timur tergolong rawan. Sementara itu, 110 kabupaten/kota lainnya dikategorikan aman. Temuan ini dapat membantu masyarakat dalam meningkatkan kewaspadaan terhadap penyebaran Covid-19.

Kemudian pada penelitian [12] menunjukkan bahwa *K-Means Clustering* efektif dalam klasifikasi sentimen Twitter, dengan *Silhouette Index* sebesar 0,687974, yang menunjukkan struktur klaster yang baik. Keberhasilan metode ini dipengaruhi oleh kualitas data latih dan uji, yang dapat memengaruhi hasil klasifikasi jika diterapkan pada dataset lain. Hasil ini sejalan dengan penelitian sebelumnya yang mendukung *K-Means Clustering* sebagai metode yang dapat diandalkan untuk analisis sentimen di media sosial. Disarankan untuk menguji metode ini pada dataset sentimen lain guna mengevaluasi konsistensi hasilnya.

Pada penelitian [13] menganalisis sentimen pengguna Twitter mengenai Pemilihan Presiden (PILPRES) 2024 di Indonesia, menggunakan algoritma *K-Means* untuk mengelompokkan 1.015 data *tweet*. Hasil menunjukkan dominasi sentimen positif (611

teks), diikuti sentimen netral (237 teks) dan negatif (89 teks). Penelitian ini memberikan wawasan penting bagi kandidat dan tim kampanye dalam merancang strategi berbasis opini publik di media sosial, serta menyoroti peran analisis sentimen dalam proses demokrasi.

Berdasarkan berbagai penelitian sebelumnya, *K-Means* merupakan metode *clustering* yang tepat untuk digunakan dalam *clustering* sentimen wisatawan di Gili, KLU, karena mampu membagi setiap kategori sentimen menjadi *cluster* yang lebih spesifik serta menghasilkan nilai *Silhouette Coefficient* untuk menilai kualitas pengelompokan. Algoritma ini dipilih karena kecepatan pemrosesannya yang tinggi, sehingga cocok untuk dataset besar dari media sosial, serta kesederhanaannya yang memungkinkan untuk dengan cepat menyusun dan menjalankan analisis. Meskipun tidak seefisien DBSCAN dalam mengidentifikasi *cluster* dengan bentuk tidak bulat, *K-Means* memberikan hasil yang baik ketika *cluster* memiliki ukuran dan kepadatan yang seragam. Berbeda dengan *Hierarchical Clustering* yang lebih lambat dan kurang efisien untuk dataset besar karena memerlukan komputasi yang lebih intensif, *K-Means* tetap sangat membantu dalam mengelompokkan data dalam bahasa Inggris dan bahasa Indonesia secara efisien. Meskipun memiliki kelemahan, seperti sensitivitas terhadap inisialisasi *centroid*, *K-Means* tetap menjadi pilihan utama dalam mengembangkan strategi yang lebih efektif untuk meningkatkan kepuasan wisatawan dan daya tarik destinasi.

2.2 Teori Penunjang

2.2.1 Analisis Sentimen

Analisis sentimen merupakan metode yang digunakan untuk menggali informasi dari teks guna mendeteksi sentimen yang dapat berupa positif, netral, atau negatif. Metode ini digunakan untuk menganalisis pandangan atau opini pengguna internet yang terpapar di media sosial, sehingga memungkinkan untuk mengidentifikasi penilaian atau pendapat pribadi yang disampaikan [14].

2.2.2 Text Mining

Text Mining adalah sebuah proses yang bertujuan untuk menghasilkan informasi yang berkualitas tinggi dari teks. Informasi yang berkualitas tinggi dapat diperoleh dengan memperhatikan pola dan tren pada teks dan menggunakan pendekatan statistik untuk mempelajarinya [15].

2.2.3 Clustering

Clustering adalah metode dalam pembelajaran mesin yang bertujuan untuk mengelompokkan data ke dalam beberapa kluster berdasarkan kesamaan karakteristik. Teknik ini digunakan untuk

mengidentifikasi pola atau struktur tersembunyi dalam data yang sebelumnya tidak diketahui [16]. Terdapat berbagai algoritma *clustering*, di antaranya *K-Means*, *Hierarchical Clustering*, dan DBSCAN. *K-Means* merupakan salah satu algoritma paling populer dan sederhana, yang membagi data ke dalam sejumlah kluster yang telah ditentukan sebelumnya. Sementara itu, *Hierarchical Clustering* membentuk hierarki kluster, sedangkan DBSCAN mengelompokkan data berdasarkan kedekatan spasial [17].

2.2.4 K-Means

K-Means merupakan metode *clustering* yang sering digunakan karena kemampuannya dalam mengelompokkan data secara efisien berdasarkan kedekatannya dengan *centroid*. Proses dimulai dengan pemilihan *centroid* awal yang ditentukan secara otomatis oleh algoritma. Selanjutnya, setiap data dihitung jaraknya ke setiap *centroid* menggunakan rumus jarak *Euclidean*, seperti Persamaan (1).

$$d(x_i, c_j) = \sqrt{\sum_{m=1}^M (x_{im} - c_{jm})^2} \quad (1)$$

Keterangan:

$d(x_i, c_j)$ = Jarak antara titik data x_i ke c_j

M = Jumlah fitur

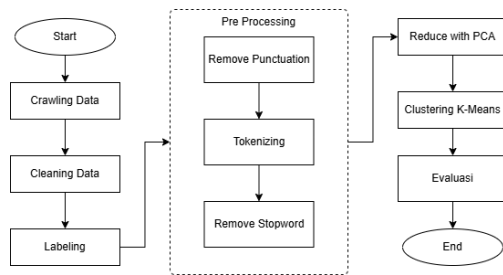
x_{im} = Nilai fitur ke- m dari titik data x_i

c_{jm} = Nilai fitur ke- m dari titik data c_j

Setelah itu, setiap data dikelompokkan ke dalam *cluster* dengan *centroid* terdekat. Setelah proses pengelompokan awal, *centroid* diperbarui dengan menghitung rata-rata dari semua titik data yang termasuk dalam *cluster* tersebut. Langkah ini terus diulang hingga perubahan posisi *centroid* menjadi minimal atau konvergen. Hasil *clustering* kemudian dievaluasi menggunakan *Silhouette Coefficient* untuk mengukur sejauh mana suatu data cocok dengan *cluster*-nya sendiri dibandingkan dengan *cluster* lain [7]. Selain itu, *K-Means* memiliki beberapa kelemahan, seperti sensitivitas terhadap pemilihan titik awal *centroid* dan ketidakmampuannya dalam mengelompokkan data dengan bentuk kluster yang tidak bulat. Meskipun demikian, algoritma ini tetap banyak digunakan karena kemudahan dalam implementasi serta kecepatan dalam memproses dan mengelompokkan data dalam skala besar [18].

3. METODE PENELITIAN

Penelitian ini dimulai dari pengumpulan data/*crawling data* dari Twitter, setelah itu dilakukan *cleaning data*, *labelling*, *preprocessing*, *reduce with PCA*, *clustering K-Means*, dan terakhir evaluasi.



Gambar 1. Flowchart Metode Penelitian

3.1 Crawling Data

Proses *crawling data* dilakukan menggunakan *tweet-harvest*, sebuah *tools* berbasis Node.js, untuk mengumpulkan *tweet* terkait Gili di KLU dalam Bahasa Indonesia dari 1 Januari 2020 hingga 5 November 2024. Data diambil dengan batasan 1000 *tweet* menggunakan autentikasi token Twitter kemudian menyimpan hasil *crawling* dalam format CSV dengan total 6030 data mentah. Tabel 1 menunjukkan hasil dari *crawling* yang masih mengandung elemen data duplikat, *@mention*, *hashtag*, emoji, URL, dan elemen lainnya.

TABEL I. CRAWLING DATA

No	Teks
1	@IShowUpdates07 mending ke NTB gili trawangan lombok lebih better, daripada ke papua agax susah.
2	di lombok ke gili trawangan cuma 15 menit semalem 400rb-an sih kt gw MANTEP BGT https://t.co/Ls8cz1Z4Gr
..	...

3.2 Cleaning Data

Data yang sudah dikumpulkan tadi kemudian akan dibersihkan dari elemen yang tidak diperlukan agar *labelling* lebih mudah dilakukan dengan membersihkan data duplikat, *@mention*, *hashtag*, emoji, URL, dan elemen lainnya. Pada tahap ini juga dilakukan penyeragaman karakter kata dengan mengubah menjadi huruf kecil, kemudian melakukan *translate* data Bahasa Indonesia ke Bahasa Inggris. Tabel 2 menunjukkan hasil dari *cleaning data*, dimana dari 6030 data hanya tersisa 3156 data.

TABEL II. CLEANING DATA

No	Teks
1	mending ke ntb gili trawangan lombok lebih better, daripada ke papua agax susah.
2	It's better to go to ntb gili trawangan lombok than to go to papua which is a bit difficult
..	...

3.3 Labelling

Pada tahap ini dilakukan *labelling* pada dataset Bahasa Indonesia menggunakan IndoNLU dan pada dataset Bahasa Inggris menggunakan DistilBERT kemudian pelabelan divalidasi secara manual oleh 3 anotator. Jika terjadi perbedaan dalam pemberian validasi pelabelan, keputusan akhir ditentukan melalui diskusi hingga mencapai kesepakatan bersama. Dengan pendekatan ini, diharapkan label sentimen yang dihasilkan tetap konsisten dan dapat diandalkan dalam analisis lebih lanjut. Tabel 3 menunjukkan hasil dari data yang telah di-*labelling*.

TABEL III. DATA LABELLING

No	Teks	Label
1	mending ke ntb gili trawangan lombok lebih better, daripada ke papua agax susah.	Positif
2	in lombok to gili trawangan it only took a few minutes last night but i was really steady	Positive
..	...	..

3.4 Pre Processing

3.4.1 Remove Punctuation

Pada tahap ini dilakukan penghapusan tanda baca yang tidak memberikan informasi yang berguna seperti titik, koma, dan tanda tanya. Tabel 4 menunjukkan hasil dari data *remove punctuation*.

TABEL IV. DATA REMOVE PUNCTUATION

No	Teks
1	mending ke ntb gili trawangan lombok lebih better daripada ke papua agax susah
2	its better to go to ntb gili trawangan lombok than to go to papua which is a bit difficult
..	...

3.4.2 Tokenizing

Pada tahap ini dilakukan pemecahan teks menjadi bagian lebih kecil untuk analisis lebih lanjut menggunakan NLTK. Tabel 5 menunjukkan hasil dari *tokenizing data*.

TABEL V. DATA TOKENIZING

No	Teks
1	['mending', 'ke', 'ntb', 'gili', 'trawangan', 'lombok', 'lebih', 'better', 'daripada', 'ke', 'papua', 'agax', 'susah']
2	['its', 'better', 'to', 'go', 'to', 'ntb', 'gili', 'trawangan', 'lombok', 'than', 'to', 'go', 'to', 'papua', 'which', 'is', 'bit', 'difficult']
..	...

3.4.3 Remove Stopword

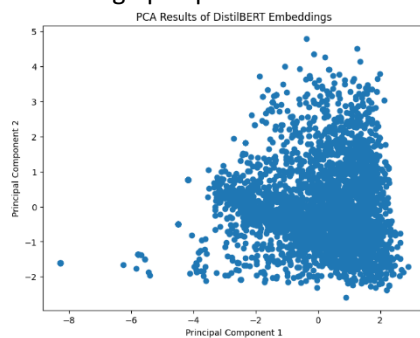
Pada tahap ini dilakukan penghapusan kata yang tidak relevan untuk memudahkan dalam analisis. Tabel 6 menunjukkan hasil dari penghapusan kata-kata yang kurang relevan pada teks.

TABEL VI. DATA REMOVE STOPWORD

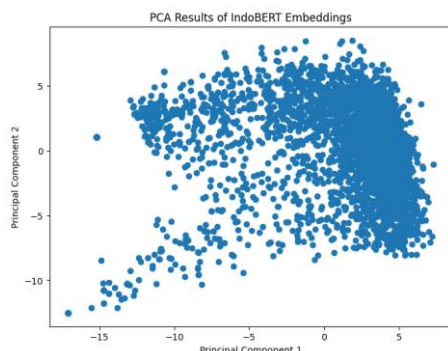
No	Teks
1	['ntb', 'gili', 'trawangan', 'lombok', 'better', 'papua', 'susah']
2	['better', 'ntb', 'gili', 'trawangan', 'lombok', 'papua', 'bit', 'difficult']
..	...

3.5 Reduce with PCA

Principal Component Analysis (PCA) digunakan untuk mengurangi dimensi data sebelum *clustering* dengan *K-Means*, mengekstrak fitur utama untuk meningkatkan efisiensi komputasi dan mengurangi *noise* tanpa kehilangan informasi penting. Dalam penelitian ini, DistilBERT dan IndoBERT digunakan karena lebih unggul dibandingkan TF-IDF dalam menangkap hubungan semantik dan nuansa emosional teks, meningkatkan akurasi sentimen serta pengelompokan data. Selain itu, model ini menghasilkan *wordcloud* yang lebih representatif, mencerminkan variasi sentimen dan isu utama secara lebih kaya, sehingga memberikan wawasan lebih mendalam tentang opini publik.



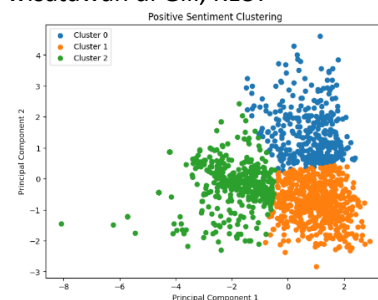
Gambar 2. Visualisasi 2D PCA DistilBERT



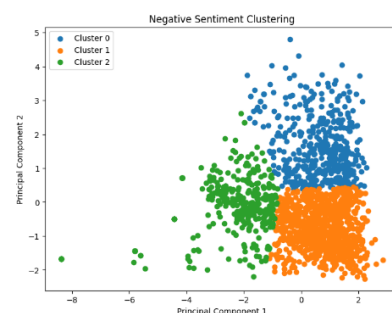
Gambar 3. Visualisasi 2D PCA IndoBERT

3.6 Clustering K-Means

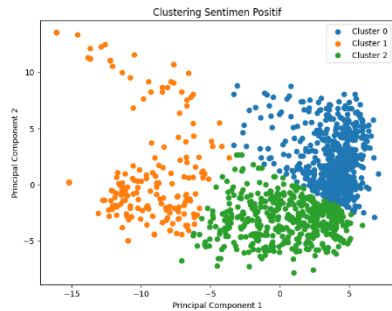
Clustering K-Means dimulai dengan mengonversi teks ulasan menjadi vektor *embedding* menggunakan DistilBERT untuk bahasa Inggris dan IndoBERT untuk bahasa Indonesia, lalu direduksi dimensinya menggunakan PCA. Proses *clustering* dilakukan berdasarkan sentimen ulasan, dengan tiga *cluster* untuk sentimen positif (*Very Positive*, *Positive*, *Quite Positive*), tiga untuk negatif (*Very Negative*, *Negative*, *Quite Negative*), dan satu untuk netral. Algoritma *K-Means* menentukan pusat *cluster* (*centroid*) secara acak, mengelompokkan data berdasarkan jarak *Euclidean* terdekat, lalu memperbarui *centroid* hingga stabil. Pemilihan $k=3$ untuk sentimen positif dan negatif bertujuan menangkap variasi intensitas opini, memungkinkan model membedakan tingkatan kepuasan dan ketidakpuasan wisatawan secara lebih rinci. Dalam dataset bahasa Indonesia, pembagian serupa diterapkan dengan label Sangat Positif, Positif, Cukup Positif serta Sangat Negatif, Negatif, Cukup Negatif, ditambah satu *cluster* untuk opini netral. Pendekatan ini meningkatkan interpretabilitas hasil, mempermudah identifikasi pola sentimen, dan memberikan wawasan lebih mendalam mengenai kepuasan wisatawan di Gili, KLU.



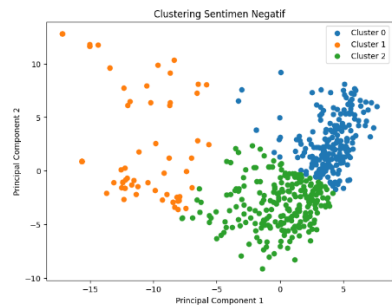
Gambar 4. Visualisasi PCA Positive Sentiment Clustering



Gambar 5. Visualisasi PCA Negative Sentiment Clustering



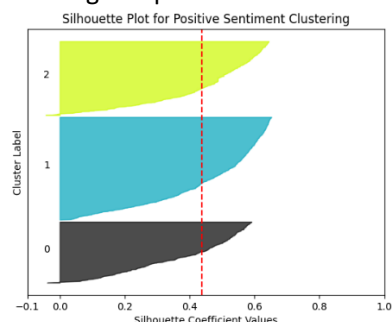
Gambar 6. Visualisasi PCA Clustering Sentimen Positif



Gambar 7. Visualisasi PCA Clustering Sentimen Negatif

3.7 Evaluasi

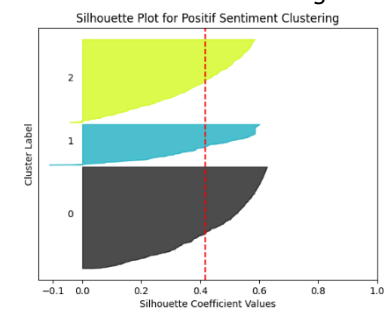
Tahap evaluasi dilakukan untuk mengukur kualitas hasil *clustering* yang terbentuk. *Silhouette Score* digunakan untuk menilai seberapa baik data dalam setiap *cluster* terpisah dari *cluster* lainnya dan seberapa dekat data dengan klasternya sendiri. Nilai *Silhouette Score* berkisar antara -1 hingga 1, dimana nilai mendekati 1 menunjukkan bahwa data dikelompokkan dengan baik, sementara nilai mendekati 0 menunjukkan bahwa data berada di perbatasan antara dua *cluster*. Selain itu, *Silhouette Score Plot* divisualisasikan untuk melihat distribusi *Silhouette Score* dari setiap data poin untuk membantu mengidentifikasi data yang salah *cluster* atau berada di perbatasan *cluster*. Analisis ini memastikan *cluster* mampu membedakan sentimen positif dan negatif dengan optimal serta memberikan gambaran yang lebih akurat mengenai pola sentimen dalam ulasan.



Gambar 8. Silhouette Plot Positive Sentiment Clustering



Gambar 9. Silhouette Plot Negative Sentiment Clustering



Gambar 10. Silhouette Plot Positif Sentiment Clustering



Gambar 11. Silhouette Plot Negatif Sentiment Clustering

4. HASIL DAN PEMBAHASAN

Model *K-Means* berhasil mengelompokkan ulasan wisatawan menjadi 3 *cluster* dari tiap sentimen berdasarkan dataset bahasa Inggris dan bahasa Indonesia dengan nilai *Silhouette Coefficient* pada sentimen positif dataset Bahasa Inggris mencapai 0,4382, sedangkan sentimen negatif 0,4258. Untuk Bahasa Indonesia mencapai 0,4173 untuk positif dan 0,4535 untuk negative. Kemudian divisualisasikan dalam bentuk *Wordcloud* untuk menampilkan kata kunci apa saja yang paling sering muncul dalam tiap *cluster*. Kemudian inspeksi manual dilakukan dengan menganalisis *full text* dari hasil *wordcloud* untuk memastikan bahwa setiap *cluster* benar-benar mencerminkan kategori sentimen yang sesuai. Analisis ini membantu mengidentifikasi potensi ambiguitas bahasa, seperti ironi atau sarkasme, yang mungkin

Pada Gambar 13 menunjukkan hasil *Wordcloud* untuk *cluster* 1 pada sentimen positif dari dataset bahasa Inggris. Berdasarkan inspeksi manual, *cluster* ini termasuk dalam kategori *positive sentiment*, yang ditandai dengan kemunculan kata-kata seperti "holiday", "good", dan "nice" yang menunjukkan bahwa wisatawan memberikan ulasan positif tentang pengalaman liburan mereka di Gili Trawangan, keindahan pantai, serta aktivitas menarik yang dapat dilakukan di sana, seperti menikmati pemandangan laut yang jernih dan matahari terbenam yang indah. Tabel 8 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.

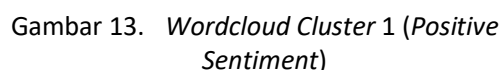


No	Teks
1	if you are on <u>holiday</u> to gili trawangan you must take the time to play subwing in the beautiful sea on the island of lombok ntb you need special techniques to be skilled at diving while being towed by a small boat gtvnews subwing news gililtrawangan ntb lombok
2	on the island of lombok gili trawangan some of the beaches here are clear and the color gradations are really <u>good</u> especially if you look directly at them i really like it
..	...

[illegible]

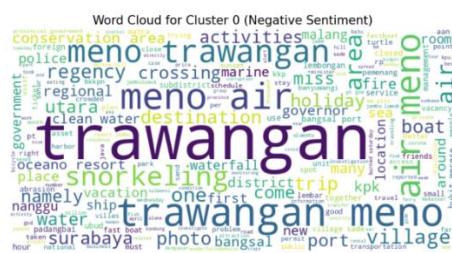
Gambar 14. *Wordcloud Cluster 2 (Positive Sentiment)*

Pada Gambar 14 menunjukkan hasil *Wordcloud* untuk *cluster* 2 pada sentimen positif dari dataset bahasa Inggris. Berdasarkan inspeksi manual, *cluster* ini termasuk dalam kategori *very positive sentiment*, yang ditandai dengan kemunculan kata-kata seperti "beautiful", "best", dan "exciting" yang menunjukkan bahwa wisatawan sangat terkesan dengan keindahan alam Gili Trawangan, Gili Meno, dan Gili Air. Ulasan mereka menekankan pesona alam yang menakjubkan, udara yang segar, dan aktivitas menarik seperti *snorkeling*, menjadikan Gili sebagai destinasi yang benar-benar memukau. Tabel 9 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.



TABEL IX. VERY POSITIVE SENTIMENT

No	Teks
1	gili trawangan gili meno and gili air are a tourist spot in lombok that offers <u>beautiful</u> natural scenery one of the tourist islands in indonesia makes this trio of gili a place full of natural wealth and amazing natural scenery
2	lombok is the <u>best</u> place ever visited especially gili trawangan my gosh its so beautiful there are no motorized vehicles there there are horses and bicycles so its so refreshing the air
..	...

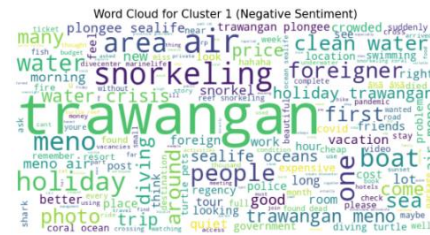


Gambar 15. Wordcloud Cluster 0 (Negative Sentiment)

Dari ilustrasi Gambar 15 menunjukkan hasil *wordcloud* untuk *cluster 0* pada sentimen negatif dari dataset Bahasa Inggris, *cluster* ini termasuk dalam kategori *quite negative sentiment*, yang ditandai dengan kemunculan kata-kata seperti "transportation", "crossing", dan "police" yang menunjukkan bahwa wisatawan mengalami kendala terkait akses transportasi menuju dan dari Gili, termasuk pembatasan perjalanan serta sistem penyeberangan yang menimbulkan polemik. Selain itu, adanya kata "police" mengindikasikan adanya isu keamanan yang menjadi perhatian wisatawan. Tabel 10 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.

TABEL X. QUITE NEGATIVE SENTIMENT

No	Teks
1	governor ntb <u>transportation</u> restrictions to gili island only from bali west nusa tenggara governor zulkieflimansyah emphasized <u>transportation</u> restrictions to the three gili tourist areas in north lombok district namely gili trawangan meno and air
2	trial of a <u>crossing</u> service with a onedoor system from and to padangbai bali port to bangsal port before entering the three gilis area in north lombok regency even though it has caused polemics among tourists
..	...

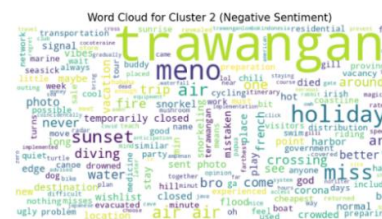


Gambar 16. Wordcloud Cluster 1 (Negative Sentiment)

Dari ilustrasi Gambar 16 menunjukkan hasil *wordcloud* untuk *cluster 1* pada sentimen negatif dari dataset Bahasa Inggris, *cluster* ini termasuk dalam kategori *negative sentiment*, yang ditandai dengan kemunculan kata-kata seperti "crisis", "expensive", dan "crowded". Kata-kata tersebut menunjukkan bahwa wisatawan menghadapi berbagai kendala, seperti krisis air bersih yang berdampak pada sektor pariwisata di Gili Trawangan, harga akomodasi dan makanan yang lebih mahal di Gili Meno dibandingkan dengan Gili lainnya, serta kondisi Gili yang sangat ramai, terutama saat musim liburan. Tabel 11 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.

TABEL XI. NEGATIVE SENTIMENT

No	Teks
1	gili trawangan tourism is dying hotels are closing due to the clean water <u>crisis</u>
2	in contrast to gili trawangan gili air and gili meno has an atmosphere that is far from the hustle and bustle following gili trawangan lodging and dining places have started to appear in gili meno only the prices are more <u>expensive</u> than the other two gili
..	...



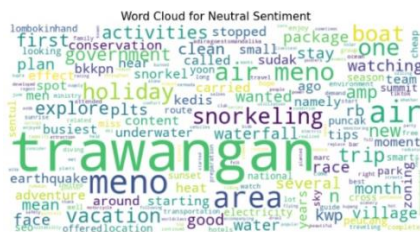
Gambar 17. Wordcloud Cluster 2 (Negative Sentiment)

Dari ilustrasi Gambar 17 menunjukkan hasil *wordcloud* untuk *cluster 0* pada sentimen negatif dari dataset Bahasa Inggris, *cluster* ini termasuk dalam kategori *very negative sentiment*, yang ditandai dengan kemunculan kata-kata seperti "fire", "evacuated", dan "temporarily closed" yang menunjukkan bahwa terdapat kejadian darurat yang berdampak signifikan pada wisatawan, seperti insiden

kebakaran speedboat di Gili Trawangan, kebakaran di salah satu hotel yang menyebabkan evakuasi wisatawan, serta penutupan sementara akses ke Gili Trawangan, Gili Air, dan Gili Meno, yang berpotensi mengganggu aktivitas wisata dan mobilitas pengunjung. Tabel 12 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.

TABEL XII. VERY NEGATIVE SENTIMENT

No	Teks
2	gill trawangan hotel north lombok ntb experienced a <u>fire</u> tourists were <u>evacuated</u>
3	access to gili trawangan air and meno are <u>temporarily closed</u>
..	...



Gambar 18. Wordcloud Neutral Sentiment

Dari ilustrasi Gambar 18 menunjukkan hasil wordcloud untuk *neutral sentiment* dari dataset Bahasa Inggris, dengan kemunculan kata-kata seperti "trip", "snorkeling", "explore", dan "snorkeling" yang mencerminkan aktivitas wisata tanpa konotasi emosional kuat. Wisatawan melakukan perjalanan ke Gili Trawangan, Gili Meno, dan Gili Air, serta menikmati snorkeling. Tabel 13 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.

TABEL XIII. NEUTRAL SENTIMENT

No	Teks
1	ever a few years ago <u>trip</u> from bali continued to gili trawangan then stopped by gili meno gili air and stayed around senggigi beach
2	summit puncak rinjani <u>snorkeling</u> in the gili trawangan area waterfall sentul area last yesterday <u>explore</u> amp
..	...

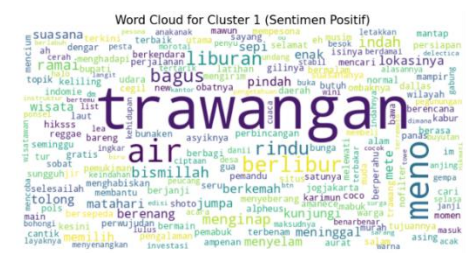


Gambar 19. Wordcloud Cluster 0 (Sentimen Positif)

Dari ilustrasi Gambar 19 menunjukkan hasil wordcloud untuk *cluster 0* pada sentimen positif dari dataset Bahasa Indonesia, *cluster* ini termasuk dalam kategori sentimen cukup positif, yang ditandai dengan kemunculan kata-kata seperti "wisata", "pantai", dan "matahari terbenam" yang mencerminkan daya tarik wisata di Gili Trawangan, seperti keberagaman destinasi, aktivitas snorkeling di Gili Meno, dan keindahan matahari terbenam di wilayah tersebut. Tabel 14 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.

TABEL XIV. SENTIMEN CUKUP POSITIF

No	Teks
1	sudah banyak tempat <u>wisata</u> di lombok tanpa mandalika banyak wisatawan mancanegara juga yang ke gili trawangan
2	gili meno lombok tidak terlalu ramai dan terlalu ramai dengan trawangan maka anda bisa snorkeling langsung dari <u>pantai</u> menuju patung sarang berjemur ini
..	...



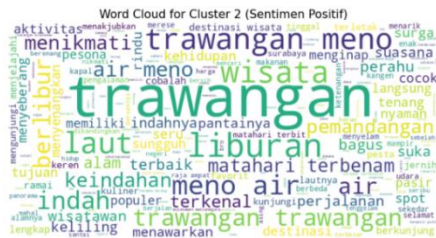
Gambar 20. Wordcloud Cluster 1 (Sentimen Positif)

Dari ilustrasi Gambar 20 menunjukkan hasil wordcloud untuk *cluster 1* pada sentimen positif dari dataset Bahasa Indonesia, *cluster* ini termasuk dalam kategori sentimen positif, yang ditandai dengan kemunculan kata-kata seperti "liburan", "bagus", dan "berlibur" yang mencerminkan antusiasme wisatawan terhadap Gili Trawangan, Gili Air, dan Gili Meno sebagai destinasi liburan yang menarik dan pilihan wisata yang direkomendasikan. Tabel 15 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.

TABEL XV. SENTIMEN POSITIF

No	Teks
1	kenapa tiba-tiba aku bermimpi lagi di lombok di mimpi seharusnya aku bekerja tapi malah terasa <u>liburan</u> sepertinya aku ingin sekali pergi ke gili trawangan gili air gili meno

2	pilihan <u>bagus</u> lombok gili trawangan raja empat nusa lembongan ubud nusa dua seribu kepulauan umalas
..	...

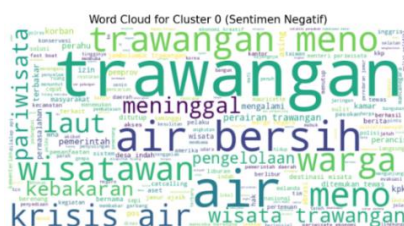


Gambar 21. *Wordcloud Cluster 2* (Sentimen Positif)

Dari ilustrasi Gambar 21 menunjukkan hasil *wordcloud* untuk *cluster 2* pada sentimen positif dari dataset Bahasa Indonesia, *cluster* ini termasuk dalam kategori sangat positif, yang ditandai dengan kemunculan kata-kata seperti "terbaik", "menikmati", dan "indah" yang menunjukkan bahwa Gili Trawangan dianggap sebagai destinasi wisata yang luar biasa, dengan pantai yang indah dan pengalaman liburan yang sangat memuaskan bagi para wisatawan. Tabel 16 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.

TABEL XVI. SENTIMEN SANGAT POSITIF

No	Teks
1	baru saja kembali ke jakarta sudah rindu gili trawangan gila <u>terbaik</u> serius gili pulau <u>terbaik</u> indonesia
2	pantai yang paling <u>indah</u> adalah gili trawangan tentunya saya tidak sempat mengambil foto karena kemarin sangat <u>menikmatinya</u> haha
3	...



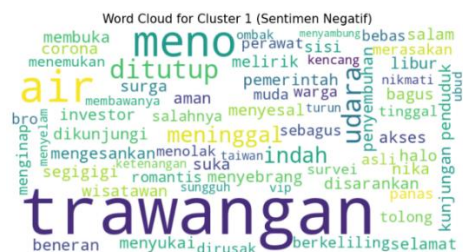
Gambar 22. *Wordcloud Cluster 0* (Sentimen Negatif)

Dari ilustrasi Gambar 22 menunjukkan hasil *wordcloud* untuk *cluster* 0 pada sentimen negatif dari dataset Bahasa Indonesia, *cluster* ini termasuk dalam kategori sangat negatif, yang ditandai dengan kemunculan kata-kata seperti "krisis air", "meninggal", dan "sulit" yang mencerminkan permasalahan serius yang dihadapi sektor pariwisata di 3 Gili tersebut

terutama terkait krisis air bersih, yang menyulitkan pelaku wisata dalam memenuhi kebutuhan dasar dan menimbulkan kekhawatiran akan menurunnya jumlah kunjungan wisatawan. Selain itu, terdapat insiden meninggalnya wisatawan di perairan sekitar Gili. Tabel 17 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.

TABEL XVII. SENTIMEN SANGAT NEGATIF

No	Teks
1	<u>Krisis air</u> bersih melanda pariwisata di gili trawangan gili meno dan gili air
2	seorang wanita paruh baya asal perancis bernama yamina belkoniene <u>meninggal</u> saat berenang di perairan gili trawangan meno dan air tramena
..	...



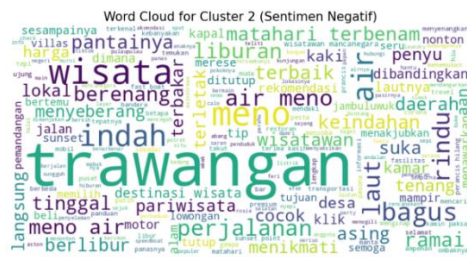
Gambar 23. *Wordcloud Cluster 1* (Sentimen Negatif)

Dari ilustrasi Gambar 23 menunjukkan hasil *wordcloud* untuk *cluster* 1 pada sentimen negatif dari dataset Bahasa Indonesia, *cluster* ini termasuk dalam kategori sentimen negatif, yang ditandai dengan kemunculan kata-kata seperti "ditutup", "akses", dan "menyesal" yang menunjukkan bahwa mencerminkan isu terkait aksesibilitas ke Gili yang ditutup akibat ombak besar serta pandangan negatif terhadap suasana di Gili, khususnya mengenai keramaian dan hiburan malam yang dianggap kurang sesuai bagi beberapa wisatawan. Tabel 18 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.

TABEL XVIII. SENTIMEN NEGATIF

No	Teks
1	lombok juga kakak saya bekerja di hotel di pulau itu katanya <u>akses</u> penyeberangan <u>ditutup</u> ombak besar seperti gili trawangan gili air gili meno dan sekitarnya
2	lebih baik jangan lakukan itu hehehehe semua penuh dengan bikini asing dan menari di malam hari lebih baik pergi ke tempat lain nanti kamu akan <u>menyesal</u> gili sudah semua meno air dan trawangan isinya sama tapi

	kalau penasaran tidak apa-apa tapi jangan menginap disana
..	...



Gambar 24. Wordcloud Cluster 2 (Negative Sentiment)

Dari ilustrasi Gambar 24 menunjukkan hasil *wordcloud* untuk *cluster* 2 pada sentimen negatif dari dataset Bahasa Indonesia, *cluster* ini termasuk dalam kategori cukup negatif, yang ditandai dengan kemunculan kata-kata seperti "perjalanan", "harga", dan "asing" yang mencerminkan keluhan wisatawan terkait harga makanan yang mahal, terutama bagi turis asing, serta berbagai kendala dalam pengelolaan perjalanan wisata di Gili, seperti keterlambatan *check-in* dan komunikasi dengan agen perjalanan. Tabel 19 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.

TABEL XIX. SENTIMEN CUKUP NEGATIF

No	Teks
1	dengan adik saya seorang konsultan <u>perjalanan</u> di lombok sering menangani tamu india di gili trawangan banyak sekali masalah late check kadang check in terlalu cepat lalu ngotot memperpanjang berbagai kerepotan lain seperti itu kadang harus telpon agennya di bali lagi
2	mampir ke pelabuhan bangsal dan tentunya berkeliling di gili air gili menyo dan yang paling terkenal adalah gili trawangan banyak turis <u>asing</u> dan berkeliling pulau dengan sepeda makanan disini mahal dan <u>harga</u> untuk turis <u>asing</u>
..	...



Gambar 25. Wordcloud Cluster 0 (Negative Sentiment)

Dari ilustrasi Gambar 25 menunjukkan hasil *wordcloud* untuk sentimen netral dari dataset Bahasa Indonesia, yang ditandai dengan kemunculan kata-kata seperti "kegiatan", "wisatawan", dan "perahu" yang mencerminkan deskripsi umum mengenai aktivitas wisata di Gili, seperti penyewaan perahu snorkeling dan imbauan kepolisian kepada wisatawan serta penyedia jasa transportasi di pelabuhan penyeberangan. Tabel 20 menunjukkan kalimat lengkap dari kata-kata yang muncul tersebut sebagai pendukung.

TABEL XX. SENTIMEN NETRAL

No	Teks
1	pelabuhan penyeberangan menuju tempat wisata gili menyo air dan trawangan dalam pelaksanaan <u>kegiatan</u> yang diadakan sat polairud polres lombok utara para personel memberikan imbauan kepada <u>wisatawan</u> dan penyedia jasa transportasi
2	menyewa <u>perahu</u> snorkeling di kawasan gili trawangan
..	...

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian mengenai analisis sentimen ulasan wisatawan terhadap destinasi Gili di KLU menggunakan *K-Means Clustering*, dapat disimpulkan bahwa terdapat variasi sentimen di antara wisatawan yang berkunjung. Sebagian besar ulasan positif menyoroti keindahan pantai, kegiatan *snorkeling*, dan suasana relaksasi, sementara ulasan negatif sering kali berkaitan dengan masalah aksesibilitas dan kondisi layanan yang perlu menjadi perhatian bagi dinas terkait untuk mengatasi permasalahan yang mengganggu kenyamanan wisatawan tersebut. Penggunaan *K-Means Clustering* dalam analisis ini berfungsi dengan cukup baik untuk mengelompokkan data sentimen, dengan hasil evaluasi yang menunjukkan nilai *Silhouette Coefficient* yang beragam. Nilai *Silhouette Coefficient* pada sentimen positif dataset Bahasa Inggris mencapai 0,4382, sedangkan sentimen negatif 0,4258. Untuk Bahasa Indonesia, nilai tersebut masing-masing 0,4173 untuk positif dan 0,4535 untuk negatif, dengan sentimen negatif memiliki nilai lebih tinggi dibandingkan Bahasa Inggris. Hal ini disebabkan oleh perbedaan struktur bahasa, kosakata, atau konteks budaya yang membuat ekspresi sentimen negatif dalam Bahasa Indonesia lebih jelas terpisah. Rentang nilai ini masih berada dalam kategori *fair*, yang menunjukkan bahwa hasil *clustering* memiliki struktur

yang cukup baik, tetapi belum optimal. Nilai *Silhouette Score* yang baik umumnya berada di atas 0,5, sehingga masih terdapat peluang untuk meningkatkan kualitas *clustering*.

Suggestion

1. Menghapus Kata yang Sering Muncul di *Wordcloud* pada Tahap *Remove Stopword*
 - Dalam penelitian ini, beberapa kata yang sering muncul dalam *wordcloud* tetap dipertahankan meskipun termasuk dalam daftar *stopwords*.
 - Ke depan, disarankan untuk menyesuaikan daftar *stopword* agar kata-kata yang tidak memiliki makna spesifik terhadap sentimen dapat dihapus, sehingga hasil *wordcloud* lebih representatif terhadap opini wisatawan.
2. Penggunaan IndoBERT Tanpa *Remove Stopword*
 - IndoBERT memiliki keunggulan dalam memahami konteks kalimat secara keseluruhan, sehingga mampu menangkap makna yang lebih mendalam.
 - Disarankan untuk tidak melakukan *remove stopwords* saat menggunakan IndoBERT, karena penghapusan kata-kata umum dapat menyebabkan hilangnya makna asli kalimat, terutama dalam bahasa Indonesia yang memiliki banyak kata penghubung dan kata bantu yang mendukung pemahaman konteks.
3. Analisis Sentimen Menggunakan *Lexicon-Based Analysis*
 - *Lexicon Analysis* adalah metode otomatis yang mengklasifikasikan sentimen berdasarkan kamus kata (*lexicon*) yang telah ditentukan.
 - Jika metode ini diterapkan, setiap kata dalam teks akan dibandingkan dengan daftar kata berpolaritas (positif, negatif, atau netral).
 - Contoh penerapannya:
 - o Jika kata "*indah, menyenangkan, luar biasa*" ditemukan dalam teks, sistem akan mengklasifikasikannya sebagai sentimen positif.
 - o Jika kata "*buruk, mahal, tidak nyaman*" muncul, maka akan dikategorikan sebagai sentimen negatif.
 - Metode ini lebih cepat dan tidak memerlukan pelabelan manual, tetapi memiliki keterbatasan dalam menangani sarkasme, konteks kalimat, dan ambiguitas makna.
 - Oleh karena itu, kombinasi antara *Lexicon Analysis* dan model berbasis pembelajaran mesin (seperti IndoBERT atau *K-Means Clustering*) dapat meningkatkan akurasi dalam analisis sentimen wisatawan.

DAFTAR PUSTAKA

- [1] Dinas Pariwisata Provinsi Nusa Tenggara Barat, "Perubahan Lombok Dinobatkan Destinasi Halal Terbaik Dunia," 2016. [Online]. Available: <http://www.disbudpar.ntbprov.go.id/perubahan-lombok-sejak-dinobatkan-destinasi-halal-terbaik-dunia>.
- [2] E. Septiani, B. Santoso, M. Mulyadi, and M. Muhdin, "Analisis Preferensi Pengunjung Kawasan Wisata Gili Meno Kabupaten Lombok Utara," *Distribusi - Journal of Management and Business*, vol. 7, no. 2, pp. 141–154, 2019.
- [3] N. Islamy, "Analisis Sektor Potensial, Dapatkah Pariwisata Menjadi Lokomotif Baru Ekonomi Nusa Tenggara Barat?," *J. Indones. Tour. Hosp. Recreat.*, vol. 2, no. 1, pp. 1–10, 2019.
- [4] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata, "Sentiment Analysis on E-Sports For Education Curriculum Using Naive Bayes and Support Vector Machine," *J. Comput. Sci. Inf.*, vol. 13, no. 2, pp. 109–122, 2020.
- [5] M. Cendana and S. D. H. Permana, "Pra-Pemrosesan Teks Pada Grup Whatsapp Untuk Pemodelan Topik," *Jurnal Mantik Penusa*, vol. 3, no. 3, pp. 107–116, 2019.
- [6] A. Ramadhan and Z. Efendi, "Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, vol. 9, pp. 18–19, 2017.
- [7] Y. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2016.
- [8] Y. W. Syaifudin and R. A. Irawan, "Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-Means," *Jurnal Informatika Polinema*, vol. 4, no. 3, pp. 189–194, 2018.
- [9] S. Ahuja and G. Dubey, "Clustering and sentiment analysis on Twitter data," 2017 2nd International Conference on Telecommunication and Networks (TEL-NET), Noida, India, pp. 1-5, 2017.
- [10] C. V. Angkoso, M. A. N. Thrisna, B. D. Satoto, and A. Kusumaningsih, "Optimasi klasifikasi sentimen menggunakan random forest dengan preprocessing K-Means clustering dan SMOTE," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 10, no. 3, pp. 389-400, 2024.
- [11] N. Nugroho and F. D. Adhinata, "Penggunaan metode K-Means dan K-Means++ sebagai clustering data Covid-19 di Pulau Jawa," *Teknika*, vol. 11, no. 3, pp. 170-179, 2022.
- [12] M. A. Z. Larasati, N. A. S. Winarsih, M. S. Rohman, and G. W. Saraswati, "Penerapan metode K-Means clustering dalam menganalisis sentimen masyarakat

- terhadap K-Popers pada Twitter," Progresif: Jurnal Ilmiah Komputer, vol. 18, no. 2, pp. 201-210, 2022.
- [13] A. S. Amin, "Analisis Sentimen Pengguna Twitter dalam Pemilihan Presiden (PILPRES) 2024 dengan Menggunakan Algoritma K-Means," Jurnal Algoritma, vol. 21, no. 1, pp. 314–321, 2024.
- [14] F. V. Sari and A. Wibowo, "Analisis sentimen pelanggan toko online Jd. Id menggunakan metode Naive Bayes classifier berbasis konversi ikon emosi," Simetris J. Tek. Mesin, Elektro dan Ilmu Komput., vol. 10, no. 2, pp. 681-686, 2019.
- [15] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, "Text mining in big data analytics," Big Data Cogn. Comput., vol. 4, no. 1, pp. 1-34, 2020.
- [16] R. W. Sari, A. Wanto, and A. P. Windarto, "Implementasi RapidMiner dengan Metode K-Means (Studi Kasus: Imunisasi Campak pada Balita Berdasarkan Provinsi)," KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer), vol. 2, no. 1, 2018.
- [17] rdiyansyah, P. A. Rahayuningsih, and R. Maulana, "Analisis Perbandingan Algoritma Klasifikasi Data Mining untuk Dataset Blogger dengan RapidMiner," Jurnal Khatulistiwa Informatika, 2018.
- [18] H. Haviluddin, S. J. Patandianan, G. M. Putra, N. Puspitasari, and H. S. Pakpahan, "Implementasi Metode K-Means untuk Pengelompokan Rekomendasi Tugas Akhir," Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer, vol. 16, no. 1, 2021.