

PENGELOMPOKAN OPINI PENGGUNA PADA KOMENTAR VIDEO YOUTUBE TENTANG BELAJAR STRATEGI MIE GACOAN MENGGUNAKAN ALGORITMA K-MEANS

Rahmat Hidayat¹, Haris Tri Saputra^{*2}, Mirdatul Husnah³, Ikhwanul Fadli⁴, M. Anjas Avin Faridas⁵,
Salma Suhaimah⁶, Juwi Retno Asih⁷, Alkas Mita⁸

¹Manajemen Informatika, FMIPA, Universitas Riau
^{2,3,4,5,6,7,8}Sistem Informasi, FMIPA, Universitas Riau

email : ¹rahmat.hidayat@lecturer.unri.ac.id,

²haristrisaputra@lecturer.unri.ac.id, ³mirdatulhusnah@lecturer.unri.ac.id,

⁴ikhwanul.fadli6222@student.unri.ac.id, ⁵m.anjas4042@student.unri.ac.id,

⁶salma.suhaimah0488@student.unri.ac.id, ⁷juwi.retno0493@student.unri.ac.id,

⁸alkas.mita1722@student.unri.ac.id

*Penulis Korespondensi

Abstrak

Penelitian ini menggunakan algoritma K-Means untuk menggabungkan pendapat pengguna dalam komentar video *YouTube* tentang strategi pemasaran Mie Gacoan. Platform seperti *YouTube* berfungsi sebagai sarana penting bagi masyarakat untuk berbagi pendapat dan informasi di era internet saat ini. Sangat penting untuk melihat komentar pelanggan untuk mengetahui bagaimana mereka melihat produk dan strategi pemasaran yang digunakan. Teknik *crawling* digunakan untuk mendapatkan data komentar sebanyak 1.846 dari akun *YouTube* @RaymondChins. Sebelum pengolahan data dilakukan, langkah-langkah *cleaning*, *case folding*, *tokenization*, *stopword removal*, dan *stemming* adalah bagian dari proses pengolahan data. Studi ini berhasil menemukan pola sentimen dalam komentar yang dibagi menjadi beberapa kelompok, masing-masing menunjukkan tanggapan positif, negatif, atau netral, dengan menggunakan algoritma *K-Means*. Hasil analisis menunjukkan bahwa algoritma ini mengungkap persepsi masyarakat terhadap Mie Gacoan secara efektif. Selain itu, mereka memberi pemasar informasi penting untuk membangun strategi yang lebih tepat sasaran. Penelitian ini tidak hanya meningkatkan pemahaman tentang pemasaran digital, tetapi juga menunjukkan bagaimana perusahaan dapat menangani persepsi konsumen di media sosial dengan menganalisis data komentar. Oleh karena itu, penelitian ini memungkinkan pengembangan lebih lanjut dengan menggunakan teknik analisis yang lebih kompleks untuk memahami lebih dalam tentang perasaan pengguna terhadap merek.

Kata kunci: algoritma *K-Means*, analisis sentimen, *YouTube*, komentar pengguna, Mie Gacoan, pemasaran digital.

GROUPING USER OPINIONS ON YOUTUBE VIDEO COMMENTS ABOUT LEARNING MIE GACOAN STRATEGY USING K-MEANS ALGORITHM

Abstract

This research uses the *K-Means* algorithm to aggregate user opinions in *YouTube* video comments about the marketing strategy of Mie Gacoan. Platforms such as *YouTube* serve as an important means for people to share opinions and information in today's internet age. It is crucial to look at customer comments to find out how they view the product, and the marketing strategies used. A *crawling* technique was used to obtain 1,846 comment data from the @RaymondChins *YouTube* account. Before data processing is performed, the steps of *cleaning*, *case folding*, *tokenization*, *stopword removal*, and *stemming* are part of the data processing process. The study successfully discovered sentiment patterns in the comments that were divided into groups, each indicating positive, negative, or neutral responses, using the *K-Means* algorithm. The analysis results show that these algorithms reveal people's perceptions of Gacoan Noodles effectively. In addition, they provide marketers with important information to build more targeted strategies. This research not only increases the understanding of digital marketing, but also shows how companies can deal with consumer perceptions on social media by analyzing comment data. Therefore, this research allows for further development by using more complex analysis techniques to understand more about users' feelings towards the brand.

Keywords: *K-Means* algorithm, sentiment analysis, *YouTube*, user comments, gacoan noodles, digital marketing

1. PENDAHULUAN

Dalam era digital yang semakin maju, *platform* berbagi video seperti *YouTube* telah berkembang menjadi sumber utama masyarakat untuk informasi dan pembelajaran. Fenomena ini menyebabkan banyak komentar di bawah video yang menunjukkan pendapat dan tanggapan pengguna tentang berbagai konten, termasuk taktik pemasaran merek terkenal seperti *Mie Gacoan*. Dalam situasi seperti ini, pemahaman publik sangat penting untuk strategi pemasaran dan pengembangan barang dan jasa yang lebih baik. Studi menunjukkan bahwa pemasar dapat menggunakan analisis sentimen komentar pengguna untuk membuat strategi yang lebih baik [1].

Objek penelitian yang dipilih ini didasarkan pada popularitas *Mie Gacoan* sebagai merek yang berhasil menarik perhatian konsumen Indonesia. *Mie Gacoan* telah menciptakan *buzz* di media sosial, terutama di *YouTube*, dengan pendekatan pemasaran yang unik dan kreatif. Komentar yang muncul di video terkait merek ini menunjukkan reaksi konsumen terhadap produk dan tren dan preferensi yang lebih luas. Oleh karena itu, pemasar dan peneliti dapat memperoleh informasi bermanfaat dari analisis komentar ini menggunakan algoritma *K-Means* [2].

Menggunakan algoritma *K-Means* untuk mengidentifikasi pola dan sentimen, variabel penelitian ini termasuk komentar pengguna sebagai data teks yang akan dikelompokkan. Keterlibatan antara variabel ini menunjukkan bagaimana tanggapan dan pendapat pengguna disusun untuk mengetahui persepsi umum tentang strategi pemasaran *Mie Gacoan*. Penelitian ini bertujuan untuk memberikan analisis persepsi publik yang lebih terorganisir dengan mengelompokkan komentar berdasarkan kesamaan konten [3][4][5].

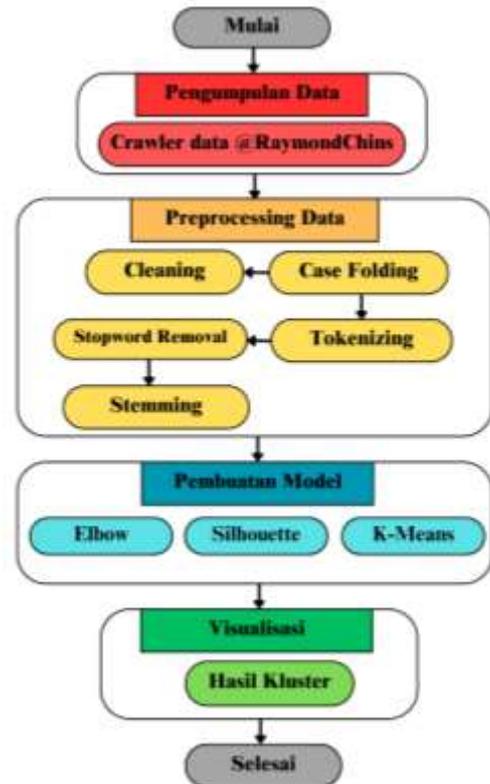
Penelitian sebelumnya telah menggunakan algoritma *K-Means* untuk analisis teks, tetapi tidak semua penelitian menggunakan metode ini untuk komentar video *YouTube* terkait merek lokal seperti *Mie Gacoan*. Studi saat ini lebih cenderung berfokus pada analisis data yang lebih umum dan tidak mempelajari bagaimana komentar pengguna dapat memberikan informasi khusus tentang strategi pemasaran merek tertentu [6].

Oleh karena itu, dengan menggunakan algoritma *K-Means* untuk menganalisis komentar pengguna secara menyeluruh, penelitian ini bertujuan untuk mengisi celah tersebut. Penemuan terbaru dalam penelitian ini adalah penggunaan analisis sentimen yang terstruktur dan komentar video *YouTube* yang berkaitan dengan *Mie Gacoan*. Manfaat dari penelitian ini adalah peningkatan pengetahuan tentang reaksi pengguna terhadap strategi pemasaran; ini dapat membantu pemasar membuat strategi yang lebih baik. Penelitian ini bertujuan untuk mendapatkan pemahaman yang lebih mendalam tentang persepsi publik terhadap merek *Mie Gacoan* dengan menerapkan algoritma

K-Means untuk mengumpulkan pendapat pengguna dan menganalisis hasilnya.

2. METODE PENELITIAN

Adapun alur dari penelitian ini seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Alur Metode Penelitian

2.1. Pengumpulan Data

Pengumpulan data dilakukan menggunakan teknik *crawling* untuk mengakses data komentar pada video *Youtube*. Data komentar berasal dari akun *Youtube @RaymondChins*. Website ini membagikan konten mengenai strategi mie pedas gacoan nomor 1 di Indonesia. Atribut data komentar yang akan kita ambil pada video tersebut, yaitu *text* (isi komentar), *author* (penulis), *likes* (jumlah like), dan *published_at* (waktu publikasi). Jumlah data komentar dari video *Youtube* yang diperoleh, yaitu 1,8 ribu baris. Pengumpulan data dilakukan dengan menggunakan bahasa pemrograman Python.

2.2. Preprocessing Data

Data akan melalui beberapa proses, tujuan proses ini adalah mengubah data mentah menjadi data bersih untuk menyederhanakan proses *clustering* [7]. Hasil akhir yang diharapkan dari *preprocessing* berupa data bersih yang siap diolah [8]. Langkah-langkah untuk *preprocessing* data meliputi:

2.2.1. Cleaning

Proses pembersihan dataset yang dilakukan bertujuan untuk menghapus tanda baca dan simbol yang tidak diperlukan dari teks [9]. Pada tahapan ini data akan dilakukan pemilahan informasi yang penting agar dapat dimanfaatkan secara baik.

2.2.2. Case Folding

Proses ini bertujuan untuk menyetarakan teks yang ada di dataset. Proses *Case Folding* pada penelitian ini adalah *lowercase*, yaitu dengan mengubah seluruh teks menjadi huruf kecil untuk konsistensi.

2.2.3. Tokenizing

Proses ini akan memecah teks menjadi unit-unit yang lebih kecil yang disebut *token*, seperti kata, frasa, atau simbol. *Tokenizing* dalam analisis sentimen membantu mengubah kata-kata terpisah untuk mendeteksi kata kunci yang relevan.

2.2.4. Stopword Removal

Proses ini akan menghilangkan kata-kata umum yang tidak penting maknanya dalam teks [7], kata-kata tersebut seperti “dan”, “yang”, “ini”, “itu”, “akan”, dan lain sebagainya.

2.2.5. Stemming

Proses ini akan mengubah kata-kata menjadi bentuk dasarnya. Dalam teks seperti komentar, sering kali terdapat variasi kata yang memiliki makna sama tetapi muncul dengan kata yang berbeda karena adanya imbuhan [10]. Misalnya, kata “berlari”, “pelari”, dan sebagainya. Sebenarnya memiliki makna inti yang sama, yaitu aktivitas “lari”, namun yang berbeda hanya kata munculnya saja. Dengan melakukan stemming, kata-kata ini diubah menjadi bentuk dasar, yaitu “lari”, sehingga makna inti dapat ditangkap dengan baik tanpa terganggu oleh imbuhan [11].

2.3. Pembuatan Model

Pemodelan merupakan tahap pemilihan model dan penerapan pemodelan dengan algoritma data mining. Tahap ini bertujuan untuk mengoptimalkan hasil penelitian. Pada penelitian ini terdapat 3 pemodelan yang dilakukan yaitu Elbow, Silhouette Coefficient, dan pembentukan K-Means. Masing-masing penjelasan model tersebut akan dibahas pada bab ini.

2.3.1. Elbow

Metode Elbow merupakan salah satu metode untuk menentukan jumlah *cluster* yang tepat melalui persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Jika nilai *cluster* pertama dengan nilai *cluster* kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka jumlah nilai *cluster* tersebut yang tepat [12]. Untuk mendapatkan perbandingannya adalah dengan

menghitung *Sum of Square Error* (SSE) dari masing-masing nilai *cluster*. Karena semakin besar jumlah nilai *cluster* K, maka nilai SSE akan semakin kecil [13]. Rumus SSE sesuai dengan persamaan.

$$SC = \sum_{k=1}^n \sum_{x_i} |x_i - c_k|^2$$

Keterangan:

K = cluster ke-c

xi = jarak data objek ke-i

ck = pusat cluster ke-i

2.3.2. Silhouette Coefficient

Silhouette Coefficient digunakan untuk melihat kualitas dan kekuatan *cluster*, seberapa baik atau buruknya suatu objek ditempatkan dalam suatu *cluster*. Metode ini merupakan gabungan dari metode separasi dan kohesi [14]. Untuk menghitung nilai *silhouette coefficient*, diperlukan perhitungan nilai *silhouette index* dari sebuah data ke-i. Nilai *silhouette coefficient* didapatkan dengan mencari nilai maksimal dari nilai *Silhouette Index Global* dari jumlah *cluster* 2 sampai jumlah *cluster* n-1, seperti pada persamaan berikut.

$$\square\square = \square\square\square\square\square\square\square\square (\square)$$

Untuk menghitung nilai SI dari sebuah data ke-i, ada 2 komponen yaitu ai dan bi. Nilai ai adalah rata-rata jarak ke-i terhadap semua data lainnya dalam satu *cluster*, sedangkan bi didapatkan dengan menghitung rata-rata jarak data ke-i terhadap semua data dari *cluster* lainnya yang tidak satu *cluster* dengan data ke-i, lalu diambil yang terkecil [14]. Berikut persamaan untuk menghitung nilai $\square\square\square$.

$$a_i^j = \frac{1}{m_j - 1} \sum_{r=1}^{m_j} d(x_i^j, x_r^j)$$

Keterangan:

j = cluster

i = index data (i = 1, 2, ..., mj) $\square\square$

\square = rata-rata jarak data ke-i terhadap semua data dalam satu *cluster*

Mj = jumlah data dalam *cluster* ke-j

$\square(\square\square\square\square, \square\square\square\square) =$ jarak data ke-i dengan data ke-r dalam satu *cluster* j. Berikut ini adalah rumus perhitungan mendapatkan nilai $\square\square\square$ dapat dilihat pada persamaan di bawah ini.

$$b_i^j = \min_{n=1..k} \left\{ \frac{1}{m_n} \sum_{r=1}^{m_n} d(x_i^j, x_r^n) \right\}$$

Keterangan:

j = cluster

i = index data (i=1,2,...mj)

$\square\square\square$ = rata-rata jarak data ke-i terhadap semua data yang tidak dalam satu *cluster* dengan data ke-i

$\square\square =$ jumlah data dalam *cluster* ke-n

$d(x_i^j, x_r^n) =$ jarak data ke-i dengan data ke-j dalam satu *cluster* n Berikut ini adalah rumus perhitungan mendapatkan nilai $\square\square\square\square$ dapat dilihat pada persamaan di bawah ini.

$$SI_i^j = \frac{b_i^j - a_i^j}{\max\{a_i^j, b_i^j\}}$$

Keterangan:

S_i^j = Silhouette Index data ke-i dalam satu cluster
 \bar{S}_i = rata-rata jarak data ke-i terhadap semua data yang tidak dalam satu cluster dengan data ke-i

\bar{S}_j = rata-rata jarak data ke-i terhadap semua data dalam satu cluster.

Berikut ini adalah rumus perhitungan mendapatkan nilai S_i^j dapat dilihat pada Persamaan 11

$$S_i^j = \frac{1}{m_j} \sum_{i=1}^{m_j} S_i^j$$

Keterangan:

\bar{S}_j = Rata-rata Silhouette Index cluster j
 S_i^j = Silhouette Index data ke-i dalam satu cluster
 m_j = jumlah data dalam cluster ke-j
 i = index data ($i = 1, 2, \dots, m_j$)

Berikut ini adalah rumus perhitungan mendapatkan nilai SI global sesuai dengan Persamaan 12.

$$SI = \frac{1}{K} \sum_{j=1}^K S_j$$

Keterangan:

\bar{S} = Rata-rata Silhouette Index dari dataset
 \bar{S}_j = Rata-rata Silhouette Index cluster j
 K = jumlah cluster

2.3.3. K-Means

Modelling akan menggunakan K-Means Clustering untuk pengelompokan komentar. K-Means Clustering adalah algoritma yang digunakan untuk membagi data ke dalam sejumlah kelompok (*cluster*) berdasarkan kemiripan atribut yang dimiliki. Pada dasarnya, algoritma ini bertujuan menemukan pusat (*centroid*) yang mewakili setiap kelompok data, dan setiap data tersebut akan dikelompokkan ke pusat terdekatnya [15]. Metode *K-Means Clustering* merupakan salah satu metode Clustering yang digunakan oleh data yang jumlahnya besar dan yang membutuhkan proses yang cepat dan efisien [7].

Berikut adalah langkah perhitungan algoritma *K-Means*:

- 1) Menentukan jumlah cluster yang diinginkan, sebut sebagai k .
- 2) Memilih k titik awal secara acak sebagai pusat cluster sementara.
- 3) Mengukur jarak data ke setiap pusat cluster, yaitu dengan rumus Euclidean Distance.

$$D(p, q) = \sqrt{(\bar{x}_{1p} - \bar{x}_{1q})^2 + (\bar{x}_{2p} - \bar{x}_{2q})^2 + \dots + (\bar{x}_{np} - \bar{x}_{nq})^2}$$

Keterangan:

$D(p, q)$ = jarak data ke-p dengan pusat cluster q
 \bar{x}_{ip} = data ke-p pada atribut data ke-r
 \bar{x}_{iq} = titik pusat ke-q pada atribut

- 4) Menempatkan data ke dalam cluster dengan pusat terdekat.

- 5) Melakukan proses iterasi dengan menentukan pusat cluster (*centroid*) baru dengan rumus pada persamaan di bawah ini.

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

Keterangan:

\bar{x}_j = centroid pada cluster
 x_{ij} = objek ke-p
 n = banyaknya objek/jumlah

- 6) Ulangi langkah 3 hingga pusat *cluster* tidak lagi berubah secara signifikan [16].

2.4. Visualisasi

Hasil klusterisasi akan disimpan di dalam library, yaitu:

- 1) Pandas
- 2) NumPy
- 3) Sklearn.feature_extraction.text.TfidfVectorizer
- 4) Nltk
- 5) matplotlib
- 6) pyplot dan seaborn

3. HASIL DAN ANALISIS

3.1. Dataset

Data yang berhasil dikumpulkan dari akun youtube @Raymond Chin adalah sebanyak 1846 komentar, data tersebut dikumpulkan dari satu video youtube yang berkaitan dengan belajar strategi Mie Gacoan.

Gambar 2. Data Set Komentar Youtube

3.2. Preprocessing Data

Pra Pemrosesan teks atau yang lebih dikenal dengan nama *text preprocessing* yaitu proses membersihkan data sebelum diolah nantinya. Pada tahapan ini terdapat 5 proses diantaranya (*Cleaning*,

Case Folding, Tokenizing, Stopword Removal, Stemming).

Gambar 3. Preprocessing Data

Hasil preprocessing ini digunakan oleh algoritma *K-Means*. Metode ini mengelompokkan komentar-komentar tersebut menjadi beberapa kelompok berdasarkan ciri-ciri yang serupa. Dalam hal ini, kelompok-kelompok ini mungkin menunjukkan pendapat yang berbeda tentang Mie Gacoan, seperti positif, negatif, netral dan diluar konteks.

```

Cluster Characteristics:
cluster  size  top_terms  sentiment_distribution
0      60  mie, beli, gaco, m  ['netral': 53, 'positif': 4, 'negatif': 1]
1      74  enak, ga, gaco, j  ['netral': 73, 'positif': 1]
2      585  mie, ga, gaco, ga  ['netral': 547, 'positif': 34, 'negatif': 4]
3      88  biasa, gaco, jui  ['netral': 83, 'positif': 4]
4      212  ayam, beli, ga, j  ['netral': 201, 'positif': 10, 'negatif': 1]

Document category distribution:
category  size
netral    422
diluar konteks  22
positif   10
negatif    4
name: count, dtype: int64
    
```

Gambar 4. Cluster Characteristics

Adapun berdasarkan gambar diatas *cluster characteristics* yang isinya itu untuk menghitung.

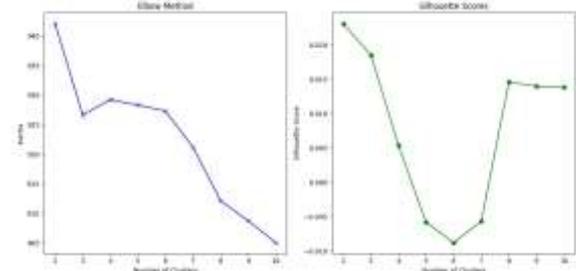
- Cluster : Nomor cluster (misalnya, 0, 1, 2, dst.) yang mengidentifikasi setiap kelompok komentar berdasarkan kesamaan kontennya.
- Size : Jumlah komentar dalam setiap cluster. Kolom ini menunjukkan seberapa besar setiap cluster, yang bisa memberikan gambaran tentang tema atau topik yang paling banyak dibahas oleh pengguna.
- Top_terms : Istilah-istilah atau kata kunci utama yang paling mewakili setiap cluster, berdasarkan perhitungan TF-IDF. Kata-kata ini memberikan gambaran singkat tentang topik utama yang dibahas dalam cluster tersebut, sehingga memudahkan untuk mengetahui tema atau isu yang banyak dibicarakan
- Sentiment_distribution : Distribusi sentimen di dalam setiap cluster. Ini biasanya berbentuk *dictionary* atau daftar, misalnya 60% positif, 30% netral, dan 10% negatif untuk suatu cluster. Kolom ini membantu memahami

sentimen umum (positif, negatif, atau netral) dari komentar yang termasuk dalam cluster tertentu.

cluster	size	top_terms	sentiment_distribution
0	60	mie, beli, gaco, m	['netral': 53, 'positif': 4, 'negatif': 1]
1	74	enak, ga, gaco, j	['netral': 73, 'positif': 1]
2	585	mie, ga, gaco, ga	['netral': 547, 'positif': 34, 'negatif': 4]
3	88	biasa, gaco, jui	['netral': 83, 'positif': 4]
4	212	ayam, beli, ga, j	['netral': 201, 'positif': 10, 'negatif': 1]

Gambar 5. Point Analisis Komentar Youtube Paling Tinggi

3.3. Pembuatan Model



Gambar 6. Hasil Elbow Method dan Silhouette Scores

Grafik Elbow Method menunjukkan penurunan inersia yang signifikan sekitar tiga hingga empat *cluster*. Setelah itu, penurunan menjadi lebih lambat, menunjukkan bahwa membentuk tiga hingga empat cluster mungkin cukup untuk mengelompokkan data dengan baik. Untuk Grafik *Silhouette Scores* menunjukkan nilai *Silhouette Scores* tertinggi sekitar 3-4 *cluster*. Ini mendukung hasil yang diperoleh dari Elbow Method.

Berdasarkan gambar hasil pengujian menggunakan elbow method maka jumlah *cluster* yang baik yang digunakan adalah 4 *cluster*, sehingga dalam penelitian ini menggunakan 4 *cluster* yaitu cluster 0, cluster 1, cluster 2, dan cluster 3.

```

Silhouette Scores by Number of Clusters:
n_clusters  silhouette_score
0            2            0.023050
1            3            0.018520
2            4            0.005344
3            5           -0.005827
4            6           -0.008783
5            7           -0.005689
6            8            0.014597
7            9            0.013991
8           10            0.013819
    
```

Gambar 7. Silhouette Scores

Silhouette score adalah metrik yang digunakan untuk mengevaluasi kualitas dari sebuah clustering. Nilai *silhouette score* berkisar antara -1 hingga 1. Semakin tinggi nilai *silhouette score*, semakin baik kualitas *clustering* yang dihasilkan. Nilai yang mendekati 1 menunjukkan bahwa suatu data point

- Komput. dan Apl.*, vol. 9, no. 02, p. 186, 2021, doi: 10.26418/coding.v9i02.47377.
- [5] F. R. Ramadhan, H. S. Wijoyo, and C. M. Saputra, "Penerapan Metode K-Means Clustering pada Ulasan Perumahan PT XYZ di Google Maps untuk Formulasi Strategi Bisnis dengan Analisis SWOT," *JPTIHK (Jurnal Pengemb. Teknol. Inf. dan Ilmu Komputer)*, vol. 7, no. 6, pp. 2879–2888, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [6] F. Alghifari Suhardi, A. Saputro, A. Sri Nugroho, and J. Heikal, "Digital Marketing Analysis of Mie Gacoan Customer at Jakarta Using RFM and K-Means Clustering Methode," *J. Econ. Bus. UBS*, vol. 13, no. 2, pp. 428–437, 2024, doi: 10.52644/joeb.v13i2.312.
- [7] N. Purnomo and W. Gata, "Pengelompokan Analisis Sentimen Komentar Youtube Terhadap Pengambilalihan Jalan Rusak di Lampung Menggunakan Algoritma Clustering," 2023.
- [8] A. R. Lahitani, "Automated Essay Scoring menggunakan Cosine Similarity pada Penilaian Esai Multi Soal," *J. Kaji. Ilm.*, vol. 22, no. 2, pp. 107–118, 2022, doi: 10.31599/jki.v22i2.1121.
- [9] K. Yolanda, Y. Yusra, and M. Fikry, "Klasifikasi Sentimen Ulasan Aplikasi WhatsApp di Play Store Menggunakan Naive Bayes Classifier," *J-Intech*, vol. 11, no. 1, pp. 1–9, 2023, doi: 10.32664/j-intech.v11i1.867.
- [10] L. H. Pramono and C. Subiyantoro, "Pengaruh Stemming Terhadap Ekstraksi Topik Menggunakan Metode Tf*Idf*Df Pada Aplikasi Pds," *JIKO (Jurnal Inform. dan Komputer)*, vol. 2, no. 1, pp. 15–23, 2017, doi: 10.26798/jiko.2017.v2i1.57.
- [11] Yeni Anistiyasari and Eko Hariadi, "Algoritma baru pembentukan kata dasar," *Pros. SNRT (Seminar Nas. Ris. Ter.)*, vol. 5662, no. November, pp. 70–76, 2019.
- [12] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.
- [13] A. Basri *et al.*, "Penentuan Jumlah Klaster Terbaik pada K-Means dalam Melihat Pola Pengelompokan Data Mahasiswa yang Telah Lulus," *JATI (Jurnal Jar. dan Teknol. Inf.)*, vol. 3, no. 1, pp. 80–86, 2023, doi: 00.0000/jati.
- [14] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
- [15] H. Al Rasyid, B. F. K. Soebari, and D. S. Y. Kartika, "Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penjualan Produk Pada Online Shop Toko Gizi," *Pros. Semin. Nas. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 242–248, 2022, doi: 10.33005/sitasi.v2i1.304.
- [16] M. Darwis, L. H. Hasibuan, M. Firmansyah, N. Ahady, and R. Tiaharyadini, "Implementation of K-Means clustering algorithm in mapping the groups of graduated or dropped-out students in the Management Department of the National University," *JISA (Jurnal Inform. dan Sains)*, vol. 4, no. 1, pp. 1–9, 2021, doi: 10.31326/jisa.v4i1.848.