

Penerapan Inset Lexicon untuk Analisis Sentimen Penonton Video JKT48 di YouTube

^{1*}Azanda Okta Kurniawan Adi, ²Ardi Sanjaya, ³Ahmad Bagus Setiawan

¹²³ Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ¹[*¹azandaadi1@gmail.com](mailto:azandaadi1@gmail.com), ²dersky@gmail.com, ³ahmadbagus@unpkediri.ac.id

Penulis Korespondens : Azanda Okta Kurniawan Adi

Abstrak— Media sosial khususnya YouTube telah menjadi platform utama interaksi antara *entertainment group* dan audiensnya, termasuk JKT48 sebagai *sister group* AKB48 di Indonesia. Penelitian ini bertujuan menganalisis sentimen komentar pada video "GO AND FIGHT! - JKT48 Special Performance Video" untuk memahami respons emosional audiens terhadap konten yang diproduksi. Metode yang digunakan adalah analisis sentimen berbasis lexicon menggunakan Inset Lexicon dengan tahapan pengambilan data melalui YouTube API v3, *text preprocessing* (*case folding, tokenizing, normalization, stopwords*, dan *stemming*), pelabelan data, dan visualisasi hasil. Dari 1000 komentar yang dianalisis, hasil menunjukkan distribusi sentimen positif 42,3%, netral 34,3%, dan negatif 21,7%. Temuan ini mengindikasikan penerimaan yang baik terhadap konten JKT48 di YouTube dan menegaskan efektivitas metode berbasis lexicon untuk analisis sentimen berbahasa Indonesia pada platform media sosial.

Kata Kunci— Analisis Sentimen, Inset Lexicon, JKT48, YouTube

Abstract— Social media, especially YouTube, has become the main platform for interaction between entertainment groups and their audiences, including JKT48 as AKB48's sister group in Indonesia. This research aims to analyze the sentiment of comments on the video "GO AND FIGHT! - JKT48 Special Performance Video" to understand the emotional response of audiences to the content produced. The method used is lexicon-based sentiment analysis using Inset Lexicon with stages of data collection through YouTube API v3, *text preprocessing* (*case folding, tokenizing, normalization, stopwords, and stemming*), data labeling, and result visualization. From 1000 comments analyzed, the results show a sentiment distribution of 42.3% positive, 34.3% neutral, and 21.7% negative. These findings indicate good acceptance of JKT48 content on YouTube and confirm the effectiveness of lexicon-based methods for Indonesian language sentiment analysis on social media platforms.

Keywords—InSet Lexicon, JKT48, Sentiment Analysis, YouTube

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Media sosial telah menjadi platform utama untuk interaksi antara *entertainment group* dan audiensnya. YouTube, sebagai platform berbagi video terbesar, memungkinkan pengguna berinteraksi melalui kolom komentar, yang menunjukkan bagaimana audiens menanggapi suatu video, memberikan kritik, dukungan, atau pandangan netral [1]. Sebagai salah satu platform video terbesar juga menyediakan ruang bagi grup idol seperti JKT48 untuk berinteraksi dengan penggemar melalui konten video dan komentar.

JKT48, sebagai sister group AKB48 di Indonesia, memiliki banyak penggemar yang aktif di platform digital. Video performance seperti "GO AND FIGHT!" merupakan salah satu konten lagu dari Setlist Oringial JKT48 yang mendapat banyak perhatian dan menghasilkan berbagai respon dari penonton. Memahami sentimen di balik komentar-komentar tersebut dapat memberikan *insight* berharga tentang *engagement* dan kepuasan audiens.

Analisis sentimen terdiri dari tiga jenis opini yaitu positif, negatif, dan netral. Analisis sentimen memahami dan mengekstraksi data opini dan secara otomatis mengolah data teks untuk mengidentifikasi sentimen yang terkandung dalam opini [2]. Analisis sentimen pada komentar YouTube menjadi semakin penting untuk memahami respons emosional audiens terhadap konten yang diproduksi.

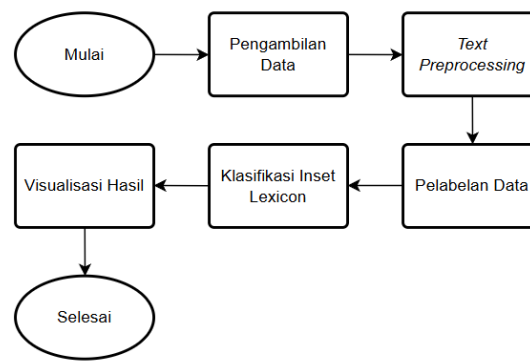
Penelitian sebelumnya dalam bidang analisis sentimen pada media sosial telah menunjukkan pentingnya pendekatan Lexicon-based untuk bahasa Indonesia. Penelitian ini menganalisis sentimen terhadap Sci-Hub. Metode Lexicon-based, dengan tahapan pengambilan data, pemuatan kamus, *pre-processing*, ekstraksi kata kunci, dan penentuan sentimen, berhasil mengklasifikasikan 59.51% sentimen positif, 26.42% negatif, dan 14.07% netral berdasarkan rule dalam kamus InSet [3]. Klasifikasi sentimen menggunakan metode Lexicon-based diawali dengan *pre-processing* untuk menentukan nilai sentimen terhadap kamus Lexicon VADER, dengan ambang batas nilai *compound* yang telah ditetapkan. Dari 2.292 komentar yang diklasifikasikan, hasilnya menunjukkan 1.788 komentar Netral, 395 Positif, dan 109 Negatif [4]. Penelitian menggunakan metode lexicon-based dengan library VADER sentiment. Berdasarkan analisis sentimen terhadap tweets, didapatkan 64% sentimen positif, 25% sentimen netral, dan 11% sentimen negatif [5]. Penelitian menggunakan 1656 data tweet dengan kata kunci yang relevan. Dengan menerapkan metode Lexicon Based, analisis sentimen berhasil mengklasifikasikan 67,2% sentimen positif, 12,6% sentimen netral, dan 20,2% sentimen negatif, yang mengindikasikan penerimaan positif dan daya tarik film tersebut di masyarakat [6]. Penelitian menganalisis opini publik di Twitter tentang kebijakan pemerintah vaksin COVID-19 menggunakan metode lexicon-based dengan library VADER sentiment. Hasilnya menunjukkan opini masyarakat cenderung netral (44,36%), dengan 23,9% agak positif, 20,25% positif, 9,6% agak negatif, dan 1,88% negatif [7].

Inset Lexicon, yang dikembangkan sebagai kamus sentimen bahasa Indonesia, telah terbukti efektif dalam berbagai aplikasi analisis sentimen. Lexicon berbasis bahasa Indonesia dapat mencapai akurasi yang tinggi dalam klasifikasi sentimen [8].

Namun, masih terdapat kesenjangan dalam penerapan analisis sentimen khusus untuk konten entertainment group Indonesia di platform YouTube. Penelitian ini bertujuan untuk mengisi gap tersebut dengan menganalisis sentimen komentar pada video JKT48 menggunakan Inset Lexicon. Tujuan penelitian ini adalah mengidentifikasi distribusi sentimen pada komentar video JKT48.

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis lexicon.



Gambar 1. Alur Penelitian

Berikut alur penelitian ini, (1) Pengambilan data dilakukan dengan *scrapping* komentar melalui YouTube API v3, (2) *Text Preprocessing*, melalui proses *case folding*, *tokenizing*, *normalization*, *stopwords* dan *stemming*, (3) Pelabelan Data, melakukan pemberian *score* berdasarkan kamus Inset Lexicon Positif dan Negatif, (4) Klasifikasi Inset Lexicon, untuk menentukan sentimen dari hasil pemberian *score*, (5) Visualisasi hasil, dilakukan untuk melihat distribusi sentimen.

A. Data Penelitian

Data yang digunakan adalah komentar-komentar pada video "GO AND FIGHT! - JKT48 Special Performance Video" yang dipublikasikan di channel YouTube resmi JKT48. Pengumpulan data dilakukan dengan proses *scrapping* melalui YouTube API v3 dengan bantuan *library googleapiclient* untuk mengambil komentar.

B. Text Preprocessing

Data komentar yang telah diambil perlu dilakukan *preprocessing*, *Preprocessing* data dalam penelitian ini melibatkan beberapa proses. Secara garis besar, tahapan-tahapan yang dilakukan adalah sebagai berikut *case folding*, *tokenizing*, *normalization*, *stopwords* dan *stemming* [9] agar proses analisis sentimen lebih akurat. *Tokenizing* dan *Stopwords* menggunakan *library Natural Language Toolkit* (NLTK). *Normalization* menggunakan dataset yang diperoleh dari Kaggle [10]. *Stemming* menggunakan *library* Sastrawi.

C. Pelabelan Inset Lexicon

Setelah data di *preprocessing*, langkah selanjutnya adalah melakukan pelabelan sentimen menggunakan Inset *Lexicon*. Inset *Lexicon* mengandung banyak kata berbahasa Indonesia yang bersifat positif dan negatif, bersama dengan bobotnya. Nilai berat kata berkisar antara -5 dan 5. Nilai minus (-) menunjukkan bahwa kata memiliki sentimen negatif, sedangkan nilai plus menunjukkan bahwa kata memiliki sentimen positif [11]. Proses ini melibatkan identifikasi kata-kata dalam kalimat yang sesuai dengan kamus *lexicon* yang berisi daftar kata positif dan negatif. Setiap kata yang teridentifikasi dalam kamus *lexicon* akan dihitung skornya berdasarkan jumlah kemunculannya dalam setiap teks atau kalimat.

$$S_{positive} = \sum_{i \in t}^n positive_score_i \quad (1)$$

$$S_{negative} = \sum_{i \in t}^n negative_score_i \quad (2)$$

Berdasarkan metode Inset Lexicon, $S_{positive}$ merepresentasikan total bobot positif suatu kalimat, yang diperoleh dari akumulasi skor polaritas setiap kata opini positif yang ditemukan di dalamnya. Demikian pula, $S_{negative}$ adalah total bobot negatif kalimat, hasil dari akumulasi skor polaritas setiap kata opini negatif.

$$Sentence_{sentiment} = \begin{cases} positif & \text{if } S_{positive} > S_{negative} \\ neutral & \text{if } S_{positive} = S_{negative} \\ negatif & \text{if } S_{positive} < S_{negative} \end{cases} \quad (3)$$

Jika jumlah kata positif dalam sebuah teks melebihi jumlah kata negatif, teks tersebut akan dikategorikan sebagai sentimen positif. Sebaliknya, jika jumlah kata positif lebih sedikit daripada kata negatif, teks akan dilabeli sebagai sentimen negatif. Apabila jumlah kata positif dan negatif dalam teks sama, maka sentimennya dianggap netral. Pelabelan Inset Lexicon menggunakan dataset yang diperoleh dari GitHub [12].

III. HASIL DAN PEMBAHASAN

A. Pengambilan Data

Pengambilan data dilakukan dengan cara *scrapping* melalui YouTube API v3 dengan bantuan library *googleapiclient* untuk mengambil komentar. Data yang digunakan adalah komentar-komentar pada video "GO AND FIGHT! - JKT48 Special Performance Video" yang dipublikasikan di channel YouTube resmi JKT48 dan mendapatkan data sebanyak 1000 Komentar. Data tersebut kemudian disimpan menggunakan format *.csv* agar memudahkan tahapan *preprocessing* data. Contoh data yang di dapat dari hasil *scrapping* dapat dilihat pada Tabel 1.

Tabel 1. Contoh Data Scrapping Komentar YouTube

<i>No</i>	<i>publishedAt</i>	<i>authorDisplayName</i>	<i>komentar</i>	<i>likeCount</i>
1	2025-06-03T01:02:12Z	@Dedeicell48	Ninggalin jejak	0
2	2025-06-02T23:35:09Z	@ekopujianto1503	Keren banget sih SPV nya, bikin salting	0
3	2025-06-02T09:40:01Z	@dayatyat8291	Suatu Saat Bertemu	1
4	2025-06-02T07:39:59Z	@krnrvhz	elin lucu bgt disini	2
5	2025-06-02T06:37:29Z	@IwanMuria	Tidak pernah BOSAN ku play lagu ini	2

B. Text Preprocessing

Pada *text preprocessing* akan mengubah data komentar yang tidak terstruktur menjadi data komentar yang terstruktur. Data komentar tersebut akan diolah dengan menggunakan metode *text mining* untuk memperoleh informasi yang baik sehingga mudah untuk dianalisis lebih lanjut. Tahapan *text preprocessing* yaitu *case folding*, *tokenizing*, *normalization*, *stopwords* dan *stemming*.

1. Case Folding

Pada proses *case folding* adalah mengubah semua huruf dalam sebuah teks menjadi huruf kecil (*lowercase*) atau terkadang menjadi huruf besar (*uppercase*), meskipun *lowercase* lebih umum.

Tabel 2. Proses Case Folding

<i>komentar</i>	<i>Case Folding</i>
Tidak pernah BOSAN ku play lagu ini	tidak pernah bosan ku play lagu ini

2. Tokenizing

Pada proses *tokenizing* adalah memecah sebuah teks menjadi unit-unit yang lebih kecil dan bermakna yang disebut token.

Tabel 3. Proses Tokenizing

<i>Case Folding</i>	<i>Tokenizing</i>
suatu saat bertemu	[suatu, saat, bertemu]

3. Normalization

Pada proses *normalization* adalah mengubah variasi bentuk kata menjadi bentuk standar.

Tabel 4. Proses Normalization

<i>Tokenizing</i>	<i>Normalization</i>
[elin, lucu, bgt, disini]	[elin, lucu, banget, disini]

4. Stopwords

Pada proses *stopwords* adalah menghapus kata-kata penghubung seperti "banget", "sih", "ya", dan "bikin".

Tabel 5. Proses Stopwords

<i>Normalization</i>	<i>Stopwords</i>
[keren, banget, sih, spv, ya, bikin, salting]	[keren, spv, salting]

5. Stemming

Pada proses *stemming* adalah Stemming adalah salah satu teknik normalisasi teks dalam *Natural Language Processing* (NLP). Tujuan utamanya adalah untuk mengurangi kata-kata berimbuhan seperti awalan, sisipan, akhiran, atau gabungan menjadi bentuk dasarnya.

Tabel 6. Proses Stemming

<i>Stopwords</i>	<i>Stemming</i>
[meninggalkan, jejak]	[tinggal, jejak]

C. Pelabelan Data

Proses selanjutnya adalah pelabelan data, agar kata dapat diolah dan dihitung. Perhitungan *score* akan dilakukan sesuai jumlah yang terdeteksi pada kamus kata Inset Lexicon. Rumus untuk menghitung *score* sentimen, yaitu dengan mengurangi jumlah kata yang memiliki *score* positif dan negatif.

$$Score = (jumlah\ kata\ positif) - (jumlah\ kata\ negatif) \quad (4)$$

Tabel 7. menunjukkan contoh hasil pelabelan data. Kata-kata yang berhasil ditemukan dalam kamus kata akan menerima *score*, yang dijumlahkan sesuai dengan rumus perhitungan *score*.

Tabel 7. Pelabelan Data

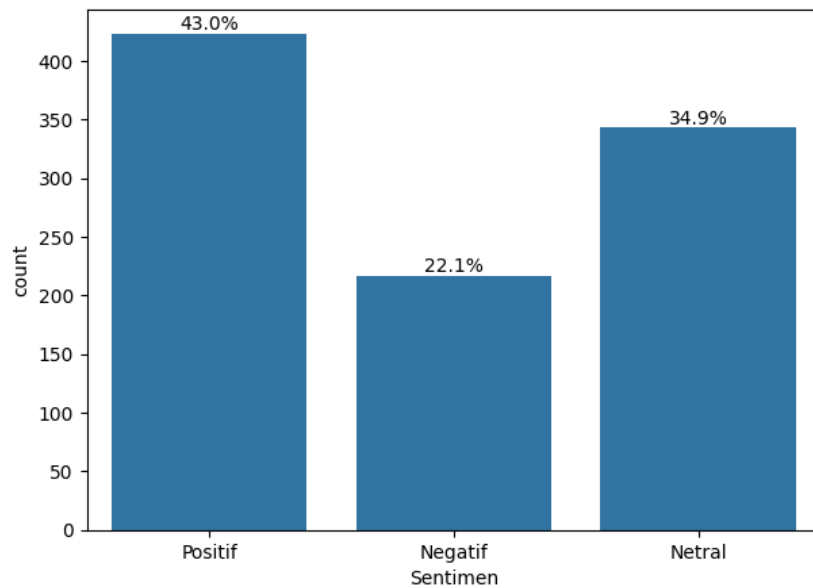
<i>Stemming</i>	<i>Score</i>	<i>Sentimen</i>
['jujur', 'gue', 'bangga', 'michie', 'center', 'mulu']	7	Positif
['jelek', 'jelek', 'persis', 'korea', 'alias', 'bening', 'sipit', 'mata']	-16	Negatif

Setelah melakukan semua proses analisis sentimen, pelabelan kelas sentimen pada 1000 data komentar untuk video "GO AND FIGHT! - JKT48 Special Performance Video" dihasilkan, dan hasilnya adalah sebagai berikut :

Tabel 8. Hasil Pelabelan Sentimen

<i>Sentimen</i>	<i>Count</i>
Positif	423
Negatif	217
Netral	343

Hasil pelabelan sentimen terhadap komentar yang dianalisis ditunjukkan dalam Tabel 8. Dari total data yang dilabeli, terdapat 423 komentar dalam kategori sentimen positif, 217 komentar dalam kategori negatif, dan 343 komentar dalam kategori netral. Hasil menunjukkan bahwa komentar dengan sentimen positif adalah yang paling banyak, disusul oleh komentar netral, dan komentar negatif adalah yang paling sedikit. Ini menunjukkan bahwa secara umum, tanggapan pengguna terhadap konten yang dianalisis cenderung positif. Gambar 2. menunjukkan visualisasi yang diperoleh dari pelabelan kelas sentimen.



Gambar 2. Visualisasi Hasil Pelabelan Sentimen

IV. KESIMPULAN

Berdasarkan hasil analisis sentimen yang telah dilakukan, maka didapatkan kesimpulan Analisis sentimen pada 1000 komentar video "GO AND FIGHT! - JKT48 Special Performance Video" menggunakan Inset Lexicon menunjukkan bahwa mayoritas sentimen audiens adalah positif (42,3%), diikuti oleh netral (34,3%) dan negatif (21,7%). Ini mengindikasikan penerimaan yang baik terhadap konten JKT48 dan menegaskan efektivitas metode berbasis lexicon untuk analisis sentimen berbahasa Indonesia di YouTube. Ke depan, penelitian ini dapat diperluas dengan menggabungkan metode berbasis *machine learning* untuk perbandingan dan validasi hasil sentimen.

DAFTAR PUSTAKA

- [1] N. Abel Laia and S. P. Barus, "Analisis Sentimen YouTube: 'Di Balik Ambisi Jokowi dalam IKN,'" *J. Pustaka AI (Pusat Akses Kaji. Teknol. Artif. Intell.*, vol. 5, no. 1, pp. 07–12, Apr. 2025, doi: 10.55382/jurnalpustakaai.v5i1.891.
- [2] M. Rizky Herdiansyah and A. Yuliana, "ANALISIS SENTIMEN KEBIJAKAN KAMPUS MERDEKA MENGGUNAKAN NAÏVE BAYES BERDASARKAN KOMENTAR PADA YOUTUBE," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 6, pp. 12454–12459, Nov. 2024, doi: 10.36040/jati.v8i6.11963.
- [3] M. Undap, V. P. Rantung, and P. T. D. Rompas, "Analisis Sentimen Situs Pembajak Artikel Penelitian Menggunakan Metode Lexicon-Based," *Jointer - J. Informatics Eng.*, vol. 2, no. 02, pp. 39–46, 2021, doi: 10.53682/jointer.v2i02.44.
- [4] S. Mujilawati, "Analisis Sentimen Pengguna Aplikasi Chatgpt Berdasarkan Rating Menggunakan Metode Lexicon," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 9, no. 1, pp. 131–137, 2023, doi: 10.36341/rabit.v9i1.3845.

- [5] S. A. Putra and A. Wijaya, "Analisis Sentimen Artificial Intelligence (Ai) Pada Media Sosial Twitter Menggunakan Metode Lexicon Based," *JuSiTik J. Sist. dan Teknol. Inf. Komun.*, vol. 7, no. 1, pp. 21–28, 2023, doi: 10.32524/jusitik.v7i1.1042.
- [6] F. Febriant, H. Christy, and A. Wijaya, "Analisis Sentimen Film The Marvels Dari Aplikasi Twitter Menggunakan Metode Lexicon Based," *JuSiTik J. Sist. dan Teknol. Inf. Komun.*, vol. 7, no. 1, pp. 37–42, 2023, doi: 10.32524/jusitik.v7i1.1046.
- [7] P. A. Sumitro, Rasiban, D. I. Mulyana, and W. Saputro, "Analisis Sentimen Terhadap Vaksin Covid-19 di Indonesia pada Twitter Menggunakan Metode Lexicon Based," *J-ICOM - J. Inform. dan Teknol. Komput.*, vol. 2, no. 2, pp. 50–56, 2021, doi: 10.33059/j-icom.v2i2.4009.
- [8] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-Janua, no. December, pp. 391–394, 2017, doi: 10.1109/IALP.2017.8300625.
- [9] S. Khomsah and A. S. Aribowo, "Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia," *J. Resti*, vol. 1, no. 3, pp. 648–654, 2017, doi: 10.13140/RG.2.2.32319.74403.
- [10] Fornigulo, "kamus_slag," [www.kaggle.com](https://www.kaggle.com/datasets/fornigulo/kamus-slag). Accessed: Jun. 04, 2025. [Online]. Available: <https://www.kaggle.com/datasets/fornigulo/kamus-slag>
- [11] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, 2021, doi: 10.57152/malcom.v1i1.20.
- [12] Fajri91, "InSet (Indonesia Sentiment Lexicon)," [github.com](https://github.com/fajri91/InSet). Accessed: Jun. 04, 2025. [Online]. Available: <https://github.com/fajri91/InSet>