

Sentimen Komentar Youtube Dengan *Sentiment Intensity Analyzer* Dari *NLTK*

Ridhwan Ardiyansyah¹, Ivan Makmur², Sabrina Phalosa Phai³

Program Studi Teknik Informatika

Universitas Tarumanagara

Jakarta Barat, Indonesia

e-mail: ¹ridhwan.535220114@stu.untar.ac.id, ²ivan.535220122@stu.untar.ac.id,

³sabrina.535220131@stu.untar.ac.id

Correspondence : e-mail: vinyam@fti.untar.ac.id

Diajukan: 30 Juni 2024; Direvisi: 12 Agustus 2024; Diterima: 25 September 2024

Abstrak

Analisis sentimen pada komentar YouTube memberikan wawasan mendalam tentang opini pengguna terhadap konten yang disajikan. Penelitian ini menggunakan *SentimentIntensityAnalyzer* dari *Natural Language Toolkit (NLTK)* untuk mengevaluasi sentimen komentar pengguna YouTube. *SentimentIntensityAnalyzer* adalah alat analisis sentimen berbasis leksikon, yang menilai polaritas sentimen dalam teks serta memberikan skor komprehensif yang mencerminkan intensitas emosi. Penelitian ini melibatkan pengumpulan komentar dari berbagai video YouTube dengan berbagai topik dan kategori. Data komentar kemudian diproses menggunakan teknik pemrosesan bahasa alami (NLP) untuk mengidentifikasi dan mengkategorikan sentimen masing-masing komentar. Proses analisis ini mencakup pra-pemrosesan data seperti penghapusan karakter khusus, tokenisasi, dan normalisasi teks untuk memastikan akurasi hasil analisis. Hasil penelitian menunjukkan bahwa *SentimentIntensityAnalyzer* mampu memberikan evaluasi yang akurat dan cepat terhadap sentimen komentar YouTube. Alat ini berhasil mengidentifikasi pola-pola sentimen yang muncul dalam berbagai kategori video. Temuan ini mengindikasikan bahwa analisis sentimen menggunakan *SentimentIntensityAnalyzer* dari *NLTK* dapat digunakan sebagai metode yang efektif untuk memahami opini publik secara luas. Dengan menggunakan analisis sentimen ini, kreator konten dan pengelola platform dapat lebih memahami reaksi audiens dan mengambil keputusan yang lebih informatif untuk meningkatkan kualitas konten dan interaksi pengguna.

Kata kunci: Analisis sentimen, YouTube, *Sentiment Intensity Analyzer*, *NLTK*.

Abstract

Sentiment analysis on YouTube comments provides deep insights into users' opinions towards the content presented. This research uses *SentimentIntensityAnalyzer* from the *Natural Language Toolkit (NLTK)* to evaluate the sentiment of YouTube user comments. *SentimentIntensityAnalyzer* is a lexicon-based sentiment analysis tool, which assesses the polarity of sentiment in text as well as provides a comprehensive score that reflects the intensity of emotion. The research involved collecting comments from various YouTube videos with various topics and categories. The comment data was then processed using natural language processing (NLP) techniques to identify and categorise the sentiment of each comment. This analysis process includes data pre-processing such as special character removal, tokenisation, and text normalisation to ensure the accuracy of the analysis results. The results show that *SentimentIntensityAnalyzer* is able to provide an accurate and quick evaluation of the sentiment of YouTube comments. The tool was able to identify sentiment patterns that appeared in different categories of videos. These findings indicate that sentiment analysis using *NLTK's SentimentIntensityAnalyzer* can be used as an effective method to understand public opinion at large. By using sentiment analysis, content creators and platform managers can better understand audience reactions and make more informed decisions to improve content quality and user interaction.

Keywords: Sentiment Analysis, YouTube, *SentimentIntensityAnalyzer*, *NLTK*

1. Pendahuluan

Di era digital saat ini, platform media sosial seperti YouTube telah menjadi sarana utama bagi individu untuk berbagi informasi, hiburan, dan edukasi. Setiap hari, jutaan pengguna YouTube meninggalkan komentar mereka di berbagai video yang mencerminkan beragam opini dan perasaan

mereka. Analisis terhadap komentar-komentar ini dapat memberikan wawasan berharga tentang bagaimana audiens merespons konten, yang pada gilirannya dapat membantu kreator konten dan pengelola *platform* dalam meningkatkan kualitas dan relevansi video yang mereka hasilkan [1]. Analisis sentimen adalah sebuah teknik untuk mendeteksi opini terhadap suatu subyek. Analisis sentimen merupakan metode untuk mengevaluasi opini dan emosi yang terkandung dalam teks. Dengan memahami sentimen yang terkandung dalam komentar YouTube, kita dapat mengidentifikasi apakah respon audiens cenderung positif, negatif, atau netral. Teknik ini berguna untuk mengukur reaksi audiens secara umum serta membantu dalam mengidentifikasi tren dan pola yang lebih spesifik dalam respon audiens terhadap berbagai jenis konten [2].

Penelitian ini menggunakan modul *SentimentIntensity Analyzer* dari *Natural Language Toolkit (NLTK)* untuk melakukan analisis sentimen pada komentar YouTube. *SentimentIntensity Analyzer* adalah alat analisis sentimen berbasis leksikon yang mampu menilai polaritas dan intensitas sentimen dalam teks [3]. Alat ini dipilih karena keandalannya dalam memberikan evaluasi sentimen yang akurat dan cepat. Penelitian ini dimulai dari pengumpulan data komentar dari berbagai topik dan kategori video YouTube. Data yang terkumpul kemudian diproses menggunakan teknik pemrosesan bahasa alami yaitu *Natural Language Processing (NLP)* untuk membersihkan dan menyiapkan teks sebelum dilakukan analisis sentimen. Hasil analisis ini dapat memberikan gambaran lebih jelas tentang pola-pola sentimen dalam komentar YouTube dan bagaimana pola-pola tersebut dapat digunakan untuk menginformasikan strategi konten dan manajemen *platform*. Dengan memahami sentimen yang diungkapkan oleh pengguna YouTube, penelitian ini dapat memberikan kontribusi signifikan dalam bidang analisis sentimen dan pengelolaan konten media sosial.

2. Metode Penelitian

Metode penelitian ini dimulai dengan pengumpulan dataset komentar dari YouTube yang beragam topik. Kemudian, data tersebut diproses dengan teknik pemrosesan bahasa alami *Natural Language Processing (NLP)* untuk membersihkan teks. Selanjutnya, kami menggunakan *SentimentIntensity Analyzer* dari *NLTK* untuk menganalisis sentimen dalam setiap komentar, menilai polaritas sentimen (positif, negatif, netral), dan memberikan skor intensitas emosi. Evaluasi akurasi dilakukan untuk memastikan ketepatan analisis [4]. Analisis pola sentimen dilakukan untuk mengetahui tren opini dalam berbagai kategori video YouTube. Hasil analisis digunakan untuk memberikan rekomendasi dalam meningkatkan kualitas konten dan interaksi pengguna di *platform*. Metode ini memungkinkan pemahaman mendalam tentang tanggapan pengguna terhadap konten YouTube. Perhitungan probabilitas prior dilakukan dengan rumus:

$$\text{Sentiment Score} = \text{Positif} - \text{Negatif} \quad (1)$$

di mana:

1. "Positif" adalah jumlah kata-kata yang memiliki sentimen positif dalam teks.
2. "Negatif" adalah jumlah kata-kata yang memiliki sentimen negatif dalam teks.

Skor sentimen yang dihasilkan adalah selisih antara jumlah kata-kata positif dan kata-kata negatif dari teks yang dianalisis. Skor positif menunjukkan sentimen positif yang dominan, skor negatif menunjukkan sentimen negatif yang dominan, dan skor netral menunjukkan keseimbangan antara sentimen positif dan negatif dalam teks [5]. *Sentiment Intensity Analyzer* juga memberikan skor intensitas emosi, yang mengukur seberapa kuat sentimen positif atau negatif yang terkandung dalam teks.

3. Hasil dan Pembahasan

3.1. Data Extraction

Data *extraction* adalah proses mengekstraksi dan mengorganisasikan data sehingga data tersebut dapat dianalisis lebih lanjut. Dengan kata lain, data *extraction* merupakan proses pengambilan informasi dari berbagai sumber untuk diolah lebih lanjut [6]. Salah satu contoh implementasinya adalah menggunakan *API* YouTube untuk mengumpulkan data komentar dari video berjudul "Deadpool & Wolverine | Main Trailer". Dengan menggunakan *YouTube API*, kita bisa mengakses data terkait video tersebut, termasuk komentar yang diposting oleh pengguna. Proses ini melibatkan beberapa langkah, mulai dari mengirim permintaan ke *API* YouTube untuk mendapatkan ID video dan komentar, hingga menguraikan respons *API* untuk mengekstrak informasi yang relevan. Hasil dari ekstraksi ini kemudian dapat disimpan dalam format yang terstruktur seperti *file CSV*, yang memungkinkan analisis lebih lanjut menggunakan alat analisis data. *File CSV* ini akan berisi kolom-kolom seperti ID komentar, nama pengguna, isi komentar, waktu komentar, dan metrik lainnya yang dapat digunakan untuk menganalisis sentimen, popularitas, dan interaksi pengguna terhadap video tersebut [7].

3.2. Data Transformation

Tahap berikutnya adalah data *transformation*. Data *Transformation* adalah proses mengubah atau mengelola data menjadi format yang lebih sesuai untuk dianalisis [8]. Pada tahap ini, komentar-komentar YouTube yang telah diekstraksi akan diolah dan disusun ke dalam format yang sesuai untuk analisis lebih lanjut. Salah satu langkah utama dalam data transformation adalah merapihkan data mentah yang sudah di *Extract* pada tahap *Data Extraction*. Pada tahap ini *library* yang digunakan yaitu:

```
#libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import os

# import functions for data preprocessing & data preparation
from sklearn.preprocessing import LabelEncoder
from sklearn.utils import resample
from sklearn.feature_extraction.text import CountVecorizer
from nltk.sentiment.vader import SentimentIntensityAnalyzer

from nltk.tokenize import word_tokenize
from nltk.stem import WordNetStemmer
from nltk.stem import PorterStemmer, LancasterStemmer
from nltk.stem.snowball import SnowballStemmer
from nltk.corpus import stopwords
from nltk.corpus import wordnet
import string
from string import punctuation
import nltk
import re
```

Gambar 1. *Library* yang digunakan

Tabel 1 menggambarkan beberapa entri komentar yang disimpan dalam *file CSV*. Setiap baris tabel mewakili satu komentar beserta informasi terkait seperti nama pengguna yang memposting komentar ("*author*"), tanggal posting komentar ("*published_at*"), jumlah komentar yang disukai ("*like_count*"), dan komentar pada video ("*text*"). Data dalam tabel ini dapat diolah lebih lanjut untuk melakukan analisis sentimen menggunakan metode *Naive Bayes*.

Tabel 1. 10 baris teratas dari *comment.csv*

	author	published_at	updated_at	like_count	Text
1	@ScreenCultureOfficial	2024-04-13T14:30:12Z	2024-04-13T14:30:12Z	28	...ey3oeWEk
2	@mrjoshuaaspeaks	2024-04-25T16:26:58Z	2024-04-25T16:26:58Z	0	...kvuNcRcOA6
3	@christiannieves5762	2024-04-25T05:24:26Z	2024-04-25T05:24:26Z	0	...wolverine fight
4	@hotrodZack1948	2024-04-24T19:00:43Z	2024-04-24T19:00:43Z	0	...rse thing to death...
5	@user-to9ux9tj8r	2024-04-24T18:13:36Z	2024-04-24T18:13:36Z	0	...APq8-AgAo">
6	@defknightssoul	2024-04-24T02:29:55Z	2024-04-24T02:29:55Z	0	...d edit, bad practice
7	@gggggglol	2024-04-23T01:05:02Z	2024-04-23T01:05:02Z	0	...VnA6aGL5M
8	@adevaughn863	2024-04-22T15:04:41Z	2024-04-22T15:04:41Z	3	...if that makes sense
9	@johnheartstone7	2024-04-22T14:51:22Z	2024-04-22T14:51:22Z	0	...no to watching this.
10	@Indigo3.7	2024-04-22T12:10:24Z	2024-04-22T12:10:24Z	0	...RIBLE PREVIEW

Setelah proses data cleaning selesai dilakukan untuk menghilangkan baris yang tidak diperlukan dalam *file CSV*, fokus akan ditempatkan pada baris komentar yang relevan. Karena adanya batasan *API* yang membatasi jumlah komentar yang dapat diekstraksi menjadi 121 baris, maka data yang tersedia akan difokuskan pada baris komentar tersebut. Selanjutnya, tabel yang berisi komentar-komentar ini akan dikelompokkan berdasarkan sentimen menjadi tiga kategori: positif, netral, dan negatif.

Tabel 2. Baris yang sudah dirapihkan

	text	Positive	Negative	Neutral	Compound	Sentiment
0	VENOM 3: THE LAST DANCE - TRAILER...	0.0	0.000	1.000	0.0000	Neutral
1	TRAILER REMIX <a href="http...	0.0	0.000	1.000	0.0000	Neutral
2	Are we getting a hulk vs wolverine fight 	0.0	0.302	0.698	-0.3818	Negative
3	Just when you thought they were done spank...	0.0	0.000	1.000	0.0000	Neutral
4	Whoa hang on, you mean everyone hasn't...	0.0	0.000	1.000	0.0000	Neutral

Fungsi *text_processing* seperti pada gambar 2, digunakan untuk membersihkan teks yang diberikan. Langkah-langkahnya meliputi konversi teks menjadi *lowercase*, penghapusan karakter *newline*, tanda baca, referensi, dan *hashtag*. Selain itu, spasi ganda, karakter khusus, dan kata-kata *stop words* juga

dihapus dari teks. Terakhir, kata-kata dalam teks diubah menjadi bentuk dasar (lemma) menggunakan *WordNetLemmatizer* dari *NLTK* [9]. Dengan demikian, teks yang telah diproses dapat digunakan.

```

1 def text_processing(text):
2     # convert text into lowercase
3     text = text.lower()
4
5     # remove new line characters in text
6     text = re.sub('\n', ' ', text)
7
8     # remove punctuations from text
9     text = re.sub('[%s]' % re.escape(punctuation), '', text)
10
11    # remove references and hashtags from text
12    text = re.sub("^a-zA-Z0-9$,.", "", text)
13
14    # remove multiple spaces from text
15    text = re.sub(r'\s+', ' ', text, flags=re.I)
16
17    # remove special characters from text
18    text = re.sub('\W', ' ', text)
19
20    text = ' '.join(word for word in word_tokenize(text) if word not in stopwords.words('english'))
21
22    # lemmatizer using WordNetLemmatizer from nltk package
23    text = ' '.join([WordNetLemmatizer().lemmatize(word) for word in word_tokenize(text)])
24
25    return text

```

Gambar 2. *text_processing*

Menggunakan *VADER SentimentIntensityAnalyzer* untuk mendapatkan skor sentimen (positif, negatif, netral, dan compound) untuk setiap teks lalu diberi label sentimen berdasarkan skor. Gambar 3 merupakan bagian dimana teks akan diberi label positif, negatif, atau netral.

```

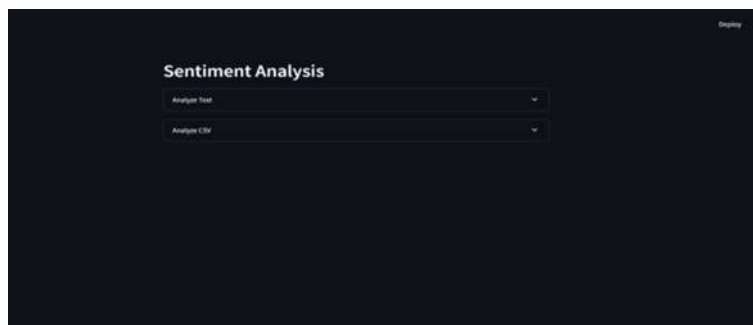
1 sentiments = SentimentIntensityAnalyzer()
2 df["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in df["text"]]
3 df["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in df["text"]]
4 df["Neutral"] = [sentiments.polarity_scores(i)["neu"] for i in df["text"]]
5 df["Compound"] = [sentiments.polarity_scores(i)["compound"] for i in df["text"]]
6 score = df["Compound"].values
7 sentiment = []
8 for i in score:
9     if i >= 0.05:
10        sentiment.append('Positive')
11    elif i <= -0.05:
12        sentiment.append('Negative')
13    else:
14        sentiment.append('Neutral')
15 df["Sentiment"] = sentiment

```

Gambar 3. *Sentiment Intensity Analyzer*

3.3. Web App

Gambar 4 menunjukkan tampilan awal dari *web app* yang dibuat menggunakan *Streamlit* untuk analisis sentimen. *Streamlit* merupakan *library* python yang berfungsi sebagai aplikasi *website* interaktif [10]. Pada tampilan awal terdapat *header* yang bertuliskan "*Sentiment Analysis*" yang memperkenalkan fungsi utama aplikasi ini. Kemudian, terdapat dua ekspander yang dapat dibuka atau ditutup oleh pengguna, yaitu "*Analyze Text*" dan "*Analyze CSV*". Ekspander "*Analyze Text*" memungkinkan pengguna untuk memasukkan dan menganalisis teks secara langsung, sementara ekspander "*Analyze CSV*" digunakan untuk mengunggah dan menganalisis *file CSV*.



Gambar 4. Tampilan Awal

4. Kesimpulan

Aplikasi *web* Streamlit yang dibuat untuk analisis sentimen ini berhasil memproses dan menganalisa komentar dari video YouTube secara efektif, menyediakan pengguna dengan alat yang kuat untuk memahami reaksi audiens terhadap konten video. Proses analisis dimulai dari ekstraksi komentar YouTube menggunakan *API*, di mana komentar dari video yang dipilih, seperti "Deadpool & Wolverine | Main Trailer," diambil dan disimpan dalam *file CSV*. Setelah *file CSV* berisi komentar diunggah ke aplikasi, proses analisis dimulai dengan memproses teks dari setiap komentar. Aplikasi ini menggunakan berbagai teknik pembersihan teks seperti menghilangkan spasi ekstra, kata-kata umum (*stopwords*), angka, dan tanda baca untuk memastikan analisis yang lebih akurat. Selain itu, aplikasi juga menyediakan fitur untuk analisis sentimen langsung dari teks yang dimasukkan oleh pengguna, menggunakan *TextBlob* untuk menghitung polaritas dan subjektivitas dari teks tersebut. Setelah pembersihan, setiap komentar dianalisis menggunakan alat analisis sentimen seperti *VADER (Valence Aware Dictionary and Sentiment Reasoner)*. *VADER* menghitung skor sentimen untuk setiap komentar, yang mencakup aspek positif, negatif, netral, dan gabungan (*compound*). Berdasarkan skor gabungan, komentar dikategorikan sebagai positif, negatif, atau netral. Hasil analisis ini kemudian ditampilkan dalam berbagai format untuk kemudahan interpretasi. Aplikasi menampilkan tabel yang mencantumkan komentar teratas berdasarkan sentimen positif, negatif, dan netral. Ini memberikan pandangan mendetail tentang bagaimana perasaan pengguna tentang konten tersebut. Selain itu, aplikasi juga menampilkan diagram batang yang menunjukkan distribusi sentimen secara keseluruhan, membantu pengguna untuk dengan cepat memahami proporsi dari berbagai jenis sentimen dalam komentar yang dianalisis. Diagram batang yang dihasilkan menyoroti jumlah komentar dalam setiap kategori sentimen, menunjukkan bahwa dalam kasus ini, komentar netral adalah yang paling banyak, diikuti oleh komentar positif, dan yang paling sedikit adalah komentar negatif. Visualisasi ini memberikan wawasan cepat dan mudah dipahami tentang persepsi audiens terhadap video tersebut. Secara keseluruhan, aplikasi ini menyediakan metode yang komprehensif dan mudah digunakan untuk melakukan analisis sentimen terhadap komentar YouTube.

Daftar Pustaka

- [1] O. Uryupina, B. Plank, A. Severyn, A. Rotondi, and A. Moschitti, "SenTube: A Corpus for Sentiment Analysis on YouTube Social Media.," in *LREC*, 2014, pp. 4244–4249.
- [2] N. M. S. Hadna, P. I. Santosa, and W. W. Winarno, "Studi literatur tentang perbandingan metode untuk proses analisis sentimen di Twitter," in *Seminar Nasional Teknologi Informasi dan Komunikasi*, 2016, pp. 57–64.
- [3] I. P. A. E. D. Udayana, I. G. A. Indrawan, and I. P. D. G. A. Putra, "Decision Support System for Sentiment Analysis of Youtube Comments on Government Policies," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 5, no. 1, pp. 27–37, Jan. 2023, doi: 10.47709/cnahpc.v5i1.1999.
- [4] J. Yao, "Automated Sentiment Analysis of Text Data with NLTK," *J Phys Conf Ser*, vol. 1187, no. 5, p. 052020, Apr. 2019, doi: 10.1088/1742-6596/1187/5/052020.
- [5] S. Elbagir and J. Yang, "Sentiment Analysis on Twitter with Python's Natural Language Toolkit and VADER Sentiment Analyzer," in *IAENG Transactions on Engineering Sciences*, WORLD SCIENTIFIC, Jan. 2020, pp. 63–80. doi: 10.1142/9789811215094_0005.
- [6] A. Darono, "Web Data Extraction Dalam Analitika Data Audit: Pengembangan Artefak Teknologi Dalam Perspektif Design Science Research," *Teknika*, vol. 9, no. 2, pp. 97–105, Nov. 2020, doi: 10.34148/teknika.v9i2.283.
- [7] G. Sanjaya and K. M. Lhaksana, "Analisis Sentimen Komentar YouTube tentang Terpilihnya Menteri Kabinet Indonesia Maju Menggunakan Lexicon Based," in *Proceedings of Engineering*, 2020.
- [8] S. Pujiono, R. Astuti, and F. Muhamad Basysyar, "Implementasi Data Mining Untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 615–620, Feb. 2024, doi: 10.36040/jati.v8i1.8360.
- [9] R. Sadida, M. R. Azkia, B. P. Candra, N. Rezeki, and M. O. C. Rendy, "Perancangan Sistem Analisis Sentimen Masyarakat Pada Sosial Media Dan Portal Berita," in *Seminar Nasional Teknologi Informasi dan Multimedia*, 2017, pp. 3–7.
- [10] J. Christian and S. I. Al Idrus, "Introduction to Citrus Fruit Ripens Using the Deep Learning Convolutional Neural Network (CNN) Learning Method," *Asian Journal of Applied Education (AJAE)*, vol. 2, no. 3, pp. 459–470, Jul. 2023, doi: 10.55927/ajae.v2i3.5003.