

# Penerapan Vader Sentiment untuk Mendeteksi Sentimen Bahasa Inggris berbasis Web

Marvin Gultom<sup>1</sup>, Jhonatan Marikros<sup>2</sup>, Wenly Rusli<sup>3</sup>

Teknik Informatika, Teknologi Informasi  
Universitas Tarumanagara  
Jakarta, Indonesia

e-mail: <sup>1</sup>marvin.535220198@stu.untar.ac.id, <sup>2</sup>jhonatan.535220189@stu.untar.ac.id,  
<sup>3</sup>wenly.535220258@stu.untar.ac.id

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

Diajukan: ....; Direvisi: ....; Diterima: ....

## Abstrak

Vader Sentiment adalah alat analisis sentimen berbasis leksikon dan aturan yang efektif untuk mendeteksi sentimen dalam teks bahasa Inggris. Penelitian ini bertujuan untuk mengimplementasikan Vader Sentiment dalam sebuah aplikasi web yang dapat mendeteksi kalimat sentimen secara real-time. Proses pengembangan aplikasi mencakup pengaturan server, pembuatan antarmuka pengguna berbasis web, dan integrasi algoritma Vader Sentiment untuk memproses teks dari pengguna. Aplikasi web ini mengkategorikan ke dalam kelas sentimen menjadi 4 kategori yaitu positif, negatif, netral, dan gabungan. Metode penelitian Lexicon Based menggunakan Library Vader Sentiment. Pada penelitian ini, juga dilakukan pendeteksian sentimen dengan mengambil 50 data komentar Twitter yang bersumber dari seorang pengguna YouTube Simplilearn. Hasil evaluasi menunjukkan bahwa akurasi deteksi sentimen aplikasi web ini mencapai 80% berdasarkan data komentar yang di uji. Penelitian ini memberikan kontribusi dalam pengembangan alat pendeteksi sentimen yang dapat digunakan secara real-time dan diakses melalui platform web.

**Kata kunci:** Lexicon Based, Pendeteksi Sentimen, Real-time, Vader Sentiment

## Abstract

Vader Sentiment is an effective lexicon and rule based sentiment analysis tool for detecting sentiment in English text. This research aims to implement Vader Sentiment in a web application that can detect sentiment sentences in real-time. The application development process includes server setup, creation of a web-based user interface, and integration of the Vader Sentiment algorithm to process text input from users. This web application categorizes sentiment classes into 4 categories, namely positive, negative, neutral and compound. The Lexicon Based research method uses the Vader Sentiment Library. In this research, sentiment detection was also carried out by taking 50 Twitter comment data sourced from a Simplilearn YouTube user. The evaluation results show that the sentiment detection accuracy of this web application reaches 80% based on the tested comment data. This research contributes to the development of a sentiment detection tool that can be used in real-time and accessed via a web platform.

**Keywords:** Lexicon Based, Real-time, Sentiment Detector, Vader Sentiment

## 1. Pendahuluan

Analisis sentimen merupakan salah satu bidang penting dalam pemrosesan bahasa alami (Natural Language Processing) yang digunakan untuk mengidentifikasi dan mengekstraksi informasi berupa pandangan (sentimen) seseorang terhadap suatu isu atau kejadian [1]. Mengingat banyaknya data teks yang dihasilkan oleh pengguna di berbagai platform media sosial, forum dan situs web, analisis sentimen sangat relevan [2]. Analisis sentimen dapat digunakan untuk mengungkap opini publik terhadap suatu isu, kepuasan pelayanan, kebijakan, dan cyber bullying. Salah satu alat yang paling populer dan efektif untuk melakukan analisis sentimen adalah Vader Sentiment, yang berbasis pada leksikon dan aturan untuk mendeteksi sentimen dalam teks berbahasa Inggris [3].

*Vader Sentiment* menawarkan keunggulan dalam mendeteksi sentimen secara akurat dengan mempertimbangkan intensitas kata-kata dan konteks penggunaannya [4]. Oleh karena itu, penelitian ini berfokus pada implementasi *Vader Sentiment* dalam sebuah aplikasi web yang mampu mendeteksi kalimat atau teks sentimen secara real-time. Tujuan utama dari penelitian ini adalah mengembangkan aplikasi website yang tidak hanya mampu memproses memasukkan teks secara manual maupun file berisi data dari pengguna dengan cepat, tetapi juga menyediakan hasil pendeteksi sentimen yang dikategorikan ke dalam tiga kelas: positif, negatif dan netral.

Penelitian ini bertujuan untuk mendeteksi sentimen teks berbahasa inggris dan mendeteksi sebuah data komentar *Twitter* yang bersumber dari pengguna youtube Simplilearn untuk mengelompokkan opini kedalam kelas polaritas yang dikategorikan ke dalam tiga kelas yaitu positif, negatif, dan netral. Setelah mengelompokkan, data tersebut akan dihitung hasil akurasi nya dengan metode *Lexicon based* menggunakan library *Vader Sentiment*.

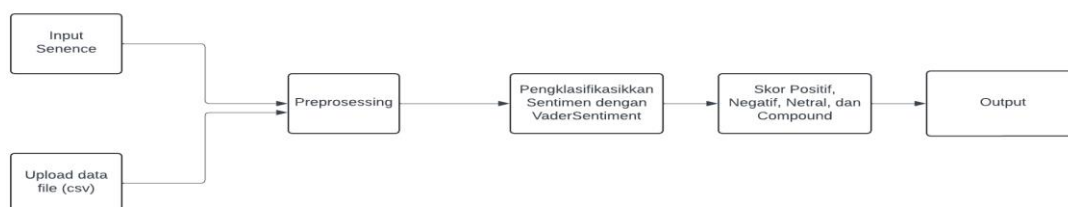
## 2. Metode Penelitian

Penelitian ini menggunakan pendekatan berbasis leksikon (*lexicon-based*) untuk analisis sentimen. *Lexicon based* adalah suatu metode atau proses pemilihan kata yang bermakna atau penting pada dokumen berdasarkan suatu kamus/leksikon yang telah tersedia. Dalam pengaplikasiannya, *Lexicon based* memiliki dua kamus yang digunakan menjadi *wordlist*. Kamus dengan kumpulan kata bersentimen positif dan kamus dengan kumpulan kata bersentimen negatif [5].

VADER (*Valance Aware Dictionary and Sentiment Reasoner*) merupakan metode *Lexicon based* dengan teknik *unsupervised* [6]. Algoritma *Vader Sentiment*, yang akan digunakan dalam penelitian ini, adalah salah satu implementasi dari pendekatan ini. *Vader Sentiment* menggunakan *leksikon* yang dirancang khusus untuk analisis sentimen dalam teks berbahasa Inggris, dan algoritma ini mempertimbangkan intensitas kata-kata serta konteks penggunaannya yang menghasilkan *class sentiment* berupa positif, negatif dan netral dengan tambahan skor total yaitu gabungan [7]. Selain itu, *Vader Sentiment* juga memiliki kamus dengan 7.500 tokens yang mengandung kata berbahasa inggris, kata *slang*, emoticons, serta sentimen yang terkait dengan singkatan, inisial dan akronim [8].

Penilaian analisis sentimen pada VADER dapat dikategorikan menjadi 4 bagian yaitu positif, negatif, netral dan gabungan. Nilai gabungan adalah nilai hasil perhitungan dengan menjumlahkan skor valensi setiap kata dalam leksikon yang telah disesuaikan dengan aturan VADER yang kemudian dinormalisasikan menjadi antara +1 (sentimen positif) dan antara -1 (sentimen negatif). Dengan adanya nilai gabungan ini, dapat digunakan sebagai aturan standar untuk mengklasifikasikan kalimat ke dalam kelas sentimen dengan ketentuan positif untuk gabungan  $\geq 0.05$ , negatif untuk gabungan  $\leq -0.05$  dan netral untuk  $-0.05 < \text{gabungan} < 0.05$  [9].

Tahap penelitian terdiri dari memberikan kalimat dan pengambilan data (*csv*), *preprocessing*, pengklasifikasian (*detection sentence*), dan output.



Gambar 2. Tahap Penelitian

Pengambilan data untuk penelitian ini berupa data komentar *twitter* yang bersumber dari pengguna *YouTube* Simplilearn. Simplilearn adalah seorang pengguna *YouTube* dengan konten-konten edukasi berupa program komputer. Data yang diambil sebanyak 50 *tweets* yang di share di *description box* di salah satu video Simplilearn. Data tersebut berupa komentar dalam bahasa inggris dengan isi komentar dengan kata-kata bersifat kasar, berjenis *slang*, dan beberapa kata-kata yang bersifat pendapat. Data tersebut diambil dan di akses oleh kami pada tanggal 15 April 2024.

*Preprocessing* adalah tahap proses awal *text mining* terhadap teks untuk mempersiapkan teks menjadi data yang dapat diolah lebih lanjut [10]. Tahap *preprocessing* memiliki tujuan membersihkan beberapa data yang siap digunakan pada tahap mendeteksi kalimat-kalimat sentimen. Proses *preprocessing* terdiri sebagai berikut:

1. *Preprocessing cleaning*: Pada proses ini dilakukan untuk menghapus tanda baca seperti “is”, “the”, “at”, dan lain-lain. Proses tersebut dinamakan proses *Stopwords*. *Stopwords* memiliki tujuan untuk memilih teks yang akan dianalisis hanya mengandung kata-kata penting untuk pemeriksaan sentimen.
2. Proses kedua menggunakan *regular expressions* (RE). Tahap ini untuk membersihkan teks dari karakter-karakter yang tidak relevan, seperti tanda baca, *URL*, tag *HTML*, *emoji*, *mention*, dan *hashtag*. Proses ini bertujuan untuk memastikan bahwa teks yang akan dianalisis hanya mengandung informasi yang penting dan relevan untuk analisis sentimen.

Setelah melakukan tahap *preprocessing*, data tersebut akan masuk kedalam tahap pengklasifikasian. Pada tahap ini dilakukan untuk melakukan *labelling sentiment* dan mengetahui *score polarity*. Metode yang digunakan adalah metode *Lexicon Based*, metode ini menggunakan kamus sebagai dasar bahasa atau leksikal. Hal ini memanfaatkan *library python* untuk *score polarity* kurang dari -0.5 adalah sentimen negatif, *score polarity* lebih besar dari 0.5 adalah sentimen positif dan *score polarity* 0 adalah sentimen netral.

Setelah semua tahap penelitian selesai, tahap selanjutnya dan tahap yang terakhir adalah tahap Output, pada tahap ini peneliti menampilkan hasil dari analisis sentimen dengan website pendeteksi sentimen berupa tabel sentiment metric dengan isi tabel adalah teks sebelum di *preprocessing*, teks setelah di *preprocessing*, score dari nilai positif, negatif, netral, gabungan, dan label yang menentukan apakah teks tersebut masuk kedalam *class sentimental* positif, negatif, atau netral.

### 3. Hasil dan Pembahasan

Pengambilan data dari youtube “Simplilearn”, data tersebut dari Twitter berbentuk csv. Data tersebut akan diuji seberapa akurat untuk mendeteksi sebuah website analisis sentimen yang kami buat. Yang dimana data akan diuji sebanyak 50 kalimat di data tersebut. Dan kami juga menguji dengan memasukkan teks langsung dari pengguna dengan bertujuan seberapa akurat website yang kami buat.

#### 3.1 Hasil Data

Pengambilan data dari youtube “Simplilearn” akan diuji sebanyak 50 kalimat dan akan dianalisis seberapa akurat dari website yang kita buat. Berikut adalah gambar tabel hasil dari 50 kalimat.

Tabel 1. Hasil data

No	Teks	teks_proses	Positif	Netral	Negatif	Gabungan	Label
0	" & you might not get ya bitch back & thats that "	amp might get ya bitch back amp thats	0.0	0.648	0.352	-0.5859	Netral
1	" @rhythmixx_ :The view from the top of the mountain was breathtaking." bitch	rhythmixx view top mountain breathtakingbitch	0.31	0.69	0.0	0.2023	Netral
2	"I am overjoyed to hear about your success. "	overjoyed hear success	0.881	0.119	0.0	0.8126	Positif
3	" So hoes that smoke are losers ? " yea ... go on IG	hoes smoke losers yea go ig	0.0	0.595	0.405	-0.5267	Netral
4	" bitch get up off me "	bitch get	0.0	0.208	0.792	-0.5859	Negatif
5	" The customer service was rude and unhelpful. "	customer service rude unhelpful	0.0	0.5	0.5	-0.4588	Netral

6	" I am so excited about the upcoming holiday season. "	excited upcoming holiday season	0.718	0.282	0.0	0.6249	Positif
7	" black bottle & a bad bitch "	black bottle amp bad bitch	0.0	0.291	0.709	-0.8074	Negatif
8	" cancel that bitch like Nino "	cancel bitch like nino	0.269	0.108	0.624	-0.5106	Negatif
9	" cant you see these hoes wont change "	cant see hoes wont change	0.0	1.0	0.0	0.0	Netral
10	" I'm feeling kinda salty about what happened. "	im feeling kinda salty happened	0.273	0.727	0.0	0.128	Netral

### 3.2 Hasil Akurasi

Dari hasil data 3.1. Bisa disimpulkan hasil data ada beberapa kalimat yang salah seharusnya masuk kedalam sentimen positif tetapi hasilnya adalah sentimen negatif. Berikut adalah contoh kalimatnya.

Tabel 2. Contoh kalimat yang salah

No	Teks	teks_proses	Positif	Netral	Negatif	Gabungan	Label
45	"I couldn't be happier with this terrible weather. " &#128517; http://t.co/R8jsymiB5b	couldn't happier terrible weather	0.348	0.273	0.379	-0.0572	Negatif

Dari contoh Tabel 2 bisa disimpulkan bahwa kalimat itu adalah kalimat negasi. Seharusnya kalimat tersebut dilabeli positif, karena terdapat kata "couldn't dan terrible" maka web kami mendeteksinya adalah kalimat negatif dimana polaritas dari kalimat itu menekankan hal positif tentang kebahagiaan terhadap cuaca buruk. Sehingga hasil akurasi dari 50 kalimat di hasil data 3.1 adalah 80% yang artinya 40 dari 50 kalimat tersebut diklasifikasikan dengan benar.

### 3.3 Hasil teks yang dimasukkan pengguna

Pada website yang kami buat selain dapat menambahkan data berupa dataset berbentuk csv, website kami juga dapat menambahkan teks yang ingin di deteksi sentimennya dari pengguna.

OMG got accepted into my dream school! yay! 🙌 #Blessed

Pilih File Tidak ada file yang dipilih

Bisa Upload CSV atau Input Langsung

Kirim

**The Sentiment of**

'omg got accepted dream school yay blessed'

**Teks is Positive!**

**Score table**

SENTIMENT METRIC	SCORE
Positive	0.887
Neutral	0.113
Negative	0.0
Compound	0.9
Average	0.33

Gambar 3. Tampilan website yang ditambahkan teks

Pada hal ini, pengguna akan memasukkan teks dari website yang kami buat. Hal ini akan menguji seberapa akurat dari website yang kami buat. Kami menguji dengan 3 hal yaitu:

1. Memasukkan menggunakan singkatan kata
2. Memasukkan menggunakan kata dan huruf berulang

## 3. Memasukkan menggunakan kata non baku

Tabel 3. Hasil memberikan teks dari pengguna

No	Jenis kata yang diuji	Teks	Hasil
1	Singkatan kata	"OMG, got accepted into my dream school! YAS! 🍀 #Blessed"	Positif
2	Kata dan huruf berulang	"i really really hate uuuuuuuuuuuuuuuuuuuuu"	Negatif
3	Kata non baku	"That movie was real dumpster fire. Total waste my time"	Negatif

Dari Tabel 3 kolom 2 pada nomor 1 bisa dilihat bahwa teks yang dimasukkan menggunakan singkatan kata bekerja dengan baik, dengan contoh kalimat yang dimasukkan adalah "OMG, got accepted into my dream school! YAS! 🍀 #Blessed", hasil dari kalimat tersebut masuk ke dalam sentimen kalimat positif. Pada kolom 3 di nomor 2 terlihat bahwa hasil dari contoh kalimat "i really really hate uuuuuuuuuuuuuuuuuuuuu" yang dimasukkan teks berupa kata dan huruf berulang. Pada website kami teks tersebut masuk ke dalam sentimen kalimat negatif. Dan pada kolom terakhir yaitu kolom 4 di nomor 3 bisa dilihat bahwa hasil teks yang dimasukkan menggunakan kata non baku dengan contoh teks "That movie was real dumpster fire. Total waste my time". Pada website kami teks tersebut masuk ke dalam sentimen kalimat negatif, menunjukkan bahwa website kami dapat bekerja dengan baik pada contoh non baku.

## 4. Kesimpulan

Penelitian ini berhasil menggunakan metode berbasis leksikon, khususnya algoritma VADER (Valence Aware Dictionary and Sentiment Reasoner), untuk mendeteksi sentimen pada komentar Twitter dari pengguna YouTube Simplilearn, dan teks yang dimasukkan oleh pengguna. Proses yang dilakukan mencakup pengambilan data, *preprocessing*, klasifikasi sentimen, dan penyajian hasil analisis. Algoritma VADER terbukti efektif dalam menganalisis sentimen teks yang mengandung bahasa kasar, emotikon, serta singkatan, inisial, dan akronim. Algoritma ini mampu mengkategorikan sentimen menjadi positif, negatif, dan netral, serta memberikan skor gabungan yang menunjukkan intensitas sentimen. Tahap *preprocessing*, yang meliputi penghapusan *stopwords* dan pembersihan teks menggunakan *regular expressions*, berhasil meningkatkan akurasi analisis sentimen dengan memastikan hanya informasi relevan yang dianalisis. Penggunaan metode berbasis leksikon dengan library Python untuk klasifikasi sentimen menunjukkan hasil yang dapat diandalkan, dengan pembagian kelas sentimen berdasarkan skor polarity dan nilai gabungan. Penyajian hasil dalam bentuk tabel metrik sentimen memberikan gambaran yang jelas tentang teks sebelum dan sesudah *preprocessing*, serta skor sentimen yang dihasilkan, sehingga mempermudah interpretasi dan evaluasi hasil analisis.

## Daftar Pustaka

- [1] A. Liawati, R. Narasati, D. Solihudin, C. L. Rohmat and S. E. Permana, "ANALISIS SENTIMEN KOMENTAR POLITIK DI MEDIA SOSIAL X DENGAN PENDEKATAN DEEP LEARNING," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, p. 7, Desember 2023.
- [2] V. Fazrian, T. Suprapti and R. Narasati, "PENERAPAN ALGORITMA NAIVE BAYES TERHADAP ANALISIS SENTIMEN APLIKASI GAME MULTIPLAYER ONLINE BATTLE ARENA ( STUDI KASUS : MOBILE LEGEND )," *JATI (Jurnal Mahasiswa Teknik Informatika)*, p. 8, 2024.
- [3] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media* 216, p. 10, 2014.
- [4] Y. Asri, W. . N. Suliyanti, D. Kuswardani and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," *PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika*, p. 12, 2022.

- 
- [5] S. Almatarneh and P. Gamallo, "A lexicon based method to search for extreme opinions," *PLOS ONE*, p. 19, 2018.
  - [6] A. H. Mulyadi and S. Lestari, "Analisis Sentimen Terhadap Sekolah Saat Covid-19 Pada Twitter Menggunakan Metode Lexicon Based," *Jurnal Informatika dan Teknologi Komputer*, vol. 3, p. 7, 2022.
  - [7] R. D. E. A, "Sentiment Analysis between VADER and EDA for the US Presidential Election 2020 on Twitter Datasets," *Journal of Applied Data Sciences*, vol. 2, p. 11, 2021.
  - [8] B. Saberi and . S. Saad, "Sentiment Analysis or Opinion Mining: A Review," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, p. 7, 2017.
  - [9] S. R. W. a. D. P. B. Y. Melani, "Analisa Kualitas Fitur Aplikasi Mobile Dengan Menggunakan Pendekatan Sentimen Grey," *Semin. Nas. Sains dan Teknol. Terap*, p. 6, 2019.
  - [10] J. Mansur, V. C. M. and T. Sutrisno, "ANALISIS PENDAPAT PUBLIK TERHADAP PUBLIC FIGURE DENGAN MENGGUNAKAN METODE NAIVE BAYES," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 7, p. 6, 2019.