

Penerapan *Vader Sentiment* untuk Mendeteksi Sentimen Bahasa Inggris berbasis *Website*

Marvin Gultom¹, Jhonatan Marikros², Wenly Rusli³

Teknik Informatika, Teknologi Informasi

Universitas Tarumanagara

Jakarta, Indonesia

e-mail: ¹marvin.535220198@stu.untar.ac.id, ²jhonatan.535220189@stu.untar.ac.id,

³wenly.535220258@stu.untar.ac.id

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

Diajukan: 24 Juni 2024; Direvisi: 12 Agustus 2024; Diterima: 18 Agustus 2024

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 website yang dapat mendeteksi kalimat sentimen secara real-time. Proses pengembangan aplikasi mencakup pengaturan server, pembuatan antarmuka pengguna berbasis website, dan integrasi algoritma *Vader Sentiment* untuk memproses teks dari pengguna. Aplikasi website ini mengkategorikan kedalam 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 website 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 website.

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 website 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 website 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 website 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 website 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 website, 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 *website* 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.

Dalam konteks metode analisis sentimen, pemilihan alat yang tepat sangat mempengaruhi hasil yang diperoleh. *Vader Sentiment*, dengan pendekatan berbasis leksikon dan aturan, menawarkan efisiensi pemrosesan yang signifikan dibandingkan dengan metode lain. Misalnya, metode berbasis pembelajaran mesin seperti *Naive Bayes* dan *Support Vector Machines (SVM)* sering kali memerlukan waktu pelatihan yang lama dan tuning parameter yang rumit. Sementara itu, model *deep learning* seperti *LSTM* dan *BERT*, meskipun akurat, memerlukan sumber daya komputasi yang tinggi dan dataset yang besar. *Vader Sentiment*, dengan kemampuannya untuk mendeteksi sentimen secara cepat tanpa memerlukan pelatihan tambahan, menjadi pilihan yang sangat sesuai untuk aplikasi *web real-time*.

Penelitian ini bertujuan untuk mendeteksi sentimen teks berbahasa Inggris dan mendeteksi sebuah data komentar *Twitter* yang bersumber dari pengguna *Youtube* “*Simplilearn*” untuk mengelompokkan opini ke dalam 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*.

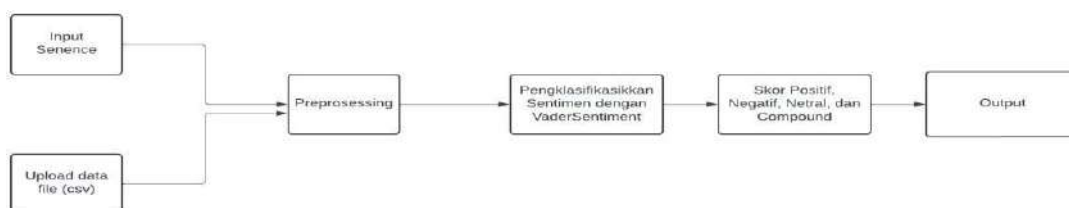
2. Metode Penelitian

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 *score* 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 *score* 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 ≥ 0.05 , negatif untuk gabungan ≤ -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 1. Tahap Penelitian
UNIVERSITAS TEKNOLOGI BANDUNG – 31 AGUSTUS 2024

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 diakses oleh kami pada tanggal 15 April 2024.

Pemilihan dataset berupa 50 *tweet* dari pengguna *YouTube* “*Simplilearn*” berpengaruh terhadap hasil analisis sentimen karena dataset ini mencakup komentar yang beragam, termasuk kata-kata kasar, *slang*, dan pendapat subjektif, yang dapat mempengaruhi akurasi dan interpretasi sentimen. Validitas dataset dapat dipertanyakan jika tidak representatif terhadap populasi komentar secara umum atau jika terdapat bias dalam pengambilan data. Oleh karena itu, penting untuk mempertimbangkan ukuran dan karakteristik dataset serta potensi dampak bias dalam hasil analisis sentimen.

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 ke dalam 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 ke dalam *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 wontchange	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. " 😅 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 website kami mendeteksinya adalah kalimat negatif dimana polaritas dari kalimat itu menekankan hal positif tentang kebahagiaan terhadap cuaca buruk. Hal ini menunjukkan kekurangan dari algoritma VADER dalam menangani negasi yang kompleks. Algoritma VADER, meskipun efektif dalam menganalisis sentimen dengan kata-kata positif dan negatif secara umum, kadang-kadang kesulitan dalam mengidentifikasi konteks negasi yang lebih nuansial. Misalnya, dalam kalimat tersebut, kata "couldn't" yang merupakan negasi dan "terrible" yang sering dianggap negatif dapat menyebabkan kesalahan dalam penilaian sentimen keseluruhan, mengabaikan konteks bahwa kalimat tersebut secara keseluruhan menyiratkan kebahagiaan. Negasi sering kali membingungkan VADER karena algoritma ini mengandalkan aturan berbasis kamus untuk menentukan polaritas kata tanpa mempertimbangkan sepenuhnya perubahan makna akibat struktur kalimat yang kompleks. Sehingga hasil akurasi dari 50 kalimat di hasil data 3.1 adalah 80% yang artinya 40 dari 50 kalimat tersebut diklasifikasikan dengan benar.

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 *score* 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 *score polarity* dan nilai gabungan. Penyajian hasil dalam bentuk tabel metrik sentimen memberikan gambaran yang jelas tentang teks sebelum dan sesudah *preprocessing*, serta *score* sentimen yang dihasilkan, sehingga mempermudah interpretasi dan evaluasi hasil analisis.

Meskipun algoritma VADER telah terbukti efektif, ada ruang untuk perbaikan dalam deteksi sentimen, terutama untuk teks yang sangat informal atau penuh *slang*. Penelitian lebih lanjut dapat fokus pada pengembangan dan integrasi model yang dapat menangani bahasa yang lebih variatif dan konteks yang lebih kompleks. Misalnya, menggabungkan algoritma berbasis leksikon dengan model pembelajaran mendalam dapat meningkatkan akurasi dalam mendeteksi nuansa sentimen yang lebih halus. Selain itu, eksplorasi terhadap teknik-teknik untuk mengurangi kesalahan deteksi pada teks yang mengandung bahasa kasar atau emotikon juga akan memberikan kontribusi berharga terhadap kemajuan dalam analisis sentimen.

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. Suprpti 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 Media216*, 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.