

Analisa Sentimen Review Produk *Handphone* Pada Situs *Amazon* Menggunakan Pendekatan *Lexicon* Berdasarkan *Sentiwordnet*

Anugerah Tri Siswanto, Rani Rotul Muhima, dan Septiyawan Rosetya Wardhana
Jurusan Teknik Informatika, Institut Teknologi Adhi Tama Surabaya

ABSTRACT

Sentiment analysis, a part of opinion mining, refers to a process to understand, extract, and process the textual data automatically to obtain information. It intends for investigating someone's opinion which then will be classified into groups of positive and negative sentiments. This research determined the word score in the document by Lexicon Sentiwordnet as the benchmark of word. Document review would pass several phases of pre-processing so that it became more structured. Next, the sentiwordnet interpretation was calculated to gain the term or word score in the document whether it belongs to positive or negative words based on the category during the process of POS tagging. After that, the researcher counted the term score summation, sentence score, text score, and system rate. The rating was then employed to investigate the quality of a certain handphone based on the user's review. In conclusion, the system being developed still indicated that it was not so good as its accuracy gained averagely 51.33% and the result of document classification earned averagely 0.26.

Article History

Received 03 – 08 – 2021
Revised 26 – 2 – 2022
Accepted 12 – 9 – 2022

Key words

Amazon,
Opinion Mining,
Product reviews
Sentiment Analysis,
Sentiwordnet.

ABSTRAK

Sentiment analysis merupakan bagian dari opinion mining, adalah proses memahami, mengesktrak dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah informasi. Dilakukan untuk mengetahui sikap seorang pembicara atau penulis yang kemudian diklasifikasikan ke dalam kelompok sentimen positif atau negatif. Pada penelitian ini penentuan nilai kata pada dokumen ditentukan menggunakan metode Lexicon Sentiwordnet sebagai tolak ukur nilai kata, dokumen review akan melalui beberapa tahapan pre-processing sehingga data dokumen review menjadi terstruktur. Selanjutnya dilakukan perhitungan sentiwordnet interpretation untuk mengetahui nilai suatu term atau kata pada dokumen untuk mengetahui kata tersebut, termasuk kata positif atau negatif berdasarkan kategorinya pada proses POS tagging. Kemudian dilakukan perhitungan term score summation, hitung nilai kalimat, hitung text score, dan hasil rating sistem, yang akan menghasilkan rating untuk mengetahui kualitas suatu produk handphone berdasarkan review pengguna. Sistem yang dikembangkan masih menunjukkan hasil yang kurang baik dengan menghasilkan hasil akurasi rata-rata sebesar 51,33%, dan rata-rata hasil klasifikasi dokumen 0,26.

PENDAHULUAN

E-commerce adalah aktivitas penyebaran, penjualan, pembelian, pemasaran produk (barang dan jasa), dengan memanfaatkan jaringan telekomunikasi seperti internet, televisi, atau jaringan komputer lainnya. *Amazon* adalah salah satu situs *E-commerce* yang menyediakan berbagai macam produk maupun jasa didalamnya. *Amazon* berperan sebagai perantara terjadinya proses jual dan beli dari produsen ke konsumen. *Amazon* adalah salah satu *vendor online* terbesar di dunia. Orang sering melihat produk dan ulasan produk sebelum membeli produk di situs *amazon* itu sendiri [1].

Sentiment analysis merupakan bagian dari *opinion mining*, adalah proses memahami, mengesktrak dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah informasi. Dilakukan untuk mengetahui sikap seorang pembicara atau penulis berkaitan dengan topik yang dibahas dari komentar yang diberikan. Pada penelitian [2], komentar-komentar suatu film dapat dijadikan standar bagus tidaknya sebuah film. Komentar dari seseorang yang dipengaruhi oleh emosional penulis (*sentiment analysis*) nantinya diklasifikasikan ke dalam kelompok sentiment positif atau negatif [3]. Komentar produk sangat diperlukan bagi perusahaan sebagai masukan dari konsumennya[4]. *Amazon* adalah salah satu situs *E-commerce* yang menyediakan berbagai macam produk maupun jasa didalamnya. *Amazon* berperan sebagai perantara terjadinya proses jual dan beli dari produsen ke konsumen. *Amazon* adalah salah satu *vendor online* terbesar di dunia. Orang sering melihat produk dan ulasan produk sebelum membeli produk di situs *amazon* itu sendiri. Tetapi ulasan tentang *amazon* tidak harus dari produk tetapi campuran produk dari ulasan produk dan ulasan layanan (terkait *amazon* atau Perusahaan Produk terkait)[1].

Berdasarkan penjelasan diatas, maka diperlukan adanya sebuah sistem yang dapat mengklasifikasikan komentar produk berbahasa inggris ke dalam hasil analisa sentimen positif atau negatif pada situs *amazon*, dan diharapkan dapat mempermudah konsumen mengetahui kualitas produk berdasarkan tanggapan dari publik dan konsumen tidak menghabiskan banyak waktu untuk membaca deskripsi teks yang panjang dalam ulasan. Oleh karena itu, pada penelitian dengan judul “Analisa sentimen pada produk *review handphone amazon* menggunakan pendekatan *lexicon* berdasarkan *sentiwordnet*” bertujuan untuk menentukan nilai sentimen pada komentar produk *amazon* ke dalam kelas positif dan negatif berdasarkan dokumen *review* pengguna.

TINJAUAN PUSTAKA

Text Mining

Text Mining adalah sebuah bidang baru dan menarik dari penelitian ilmu komputer yang mencoba untuk memecahkan krisis informasi yang berlebihan dengan menggabungkan teknik dari *Data Mining*, *Machine Learning*, *Natural Language Processing*, *Information Retrieval*, dan manajemen ilmu pengetahuan [5].

Klasifikasi Sentimen

Klasifikasi sentimen adalah kegiatan penambangan opini yang berkaitan dengan orientasi sentimen keseluruhan dari opini yang terkandung dalam dokumen yang diberikan. Orientasi opini dapat diklasifikasikan sebagai opini yang bersifat menentang positif atau negatif, tanggapan positif atau negatif mengenai suatu produk, opini yang menguntungkan atau tidak menguntungkan tentang suatu topik, atau sebuah peringkat terhadap sebuah opini yang muncul, sebagai contoh pada ulasan film dengan tanggapan dari penonton yang berkisar dari satu hingga lima bintang [6].

Opini Leksikon

Opini Leksikon adalah sumber daya yang mengasosiasikan orientasi sentimen dan kata-kata. Penggunaannya dalam penelitian *opinion mining* berasal dari hipotesis bahwa setiap kata dapat dianggap sebagai unit informasi opini, dan oleh karena itu dapat memberikan petunjuk untuk mendokumentasikan sentimen dan subjektivitas [6]. Pendekatan leksikon sendiri ada beberapa cara, antara lain pendekatan leksikon berbasis kamus dan pendekatan leksikon berbasis korpus.

Pendekatan leksikon berbasis kamus merupakan teknik sederhana dalam pendekatan yang didasarkan pada penggunaan sekelompok kecil kata opini dan kamus online, seperti *WordNet* atau *Thesaurus*. Metode ini mengumpulkan secara manual sekumpulan kecil kata-kata opini dengan fokus yang diketahui, dan kemudian memperluas koleksi itu dengan mencari di *WordNet* atau *Thesaurus* untuk sinonim dan antonimnya [7].

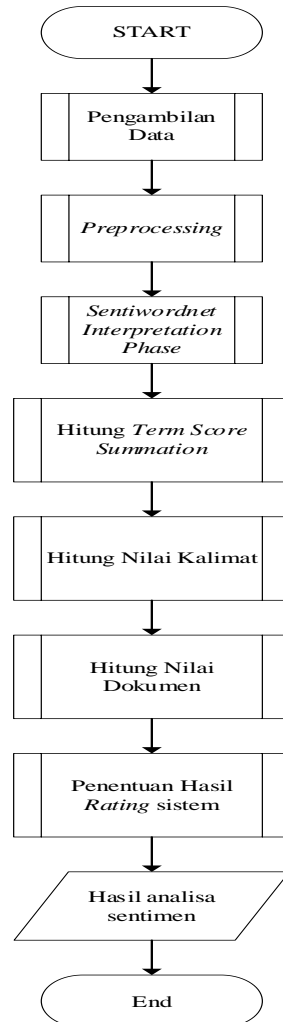
Pendekatan leksikon berbasis korpus mengandalkan sintaks atau model *co-occurrence* serta daftar kata opini untuk menemukan kata-kata opini lain dalam korpus besar. Teknik ini dimulai dengan daftar kata sifat opini dan menggunakannya bersama dengan serangkaian batasan linguistik atau konvensi konjungsi untuk mengidentifikasi kata-kata opini tambahan dari kata sifat dan disambiguasinya.

Sentiwordnet

SentiWordNet adalah sumber informasi sentimental untuk istilah dalam bahasa Inggris yang dirancang untuk membantu dalam tugas penambangan opini, di mana setiap istilah dikaitkan dengan skor numerik untuk informasi polaritas positif dan negatif. *SentiwordNet* bertujuan untuk memberikan informasi tingkat istilah tentang polaritas sebuah pendapat dengan memperoleh informasi dari basis data Bahasa Inggris *WordNet*. Untuk setiap kata di *WordNet*, skor positif dan negatif mulai dari 0 hingga 1 ada pada *SentiwordNet*. Yang menunjukkan polaritasnya, skor yang lebih tinggi menunjukkan istilah yang membawa informasi bias opini berat, sedangkan skor yang lebih rendah menunjukkan istilah yang kurang subyektif [6].

METODE

Pada penelitian ini data diambil dari komentar produk *handphone* pada situs *amazon*. Dataset yang digunakan berjumlah 5 dokumen dengan setiap dokumennya memiliki 30 komentar dengan 15 komentar positif dan 15 komentar negatif dalam bentuk csv. Secara umum flowchart sistem analisa sentimen pada penelitian digambarkan pada Gambar 1.



Gambar 1. Flowchart Sistem Analisa Sentimen

Setelah tahap pengambilan data, tahapan - tahapan analisa sentimen dengan pendekatan leksikon berbasis *sentiwordnet*, yaitu:

1. *Preprocessing*

adalah proses pertama dalam pengolahan data teks yang dilakukan untuk menghindari data yang tidak sempurna, gangguan pada data, dan data-data yang tidak konsisten [8]. Pada tahap ini terdiri dari beberapa proses, antara lain: proses penghapusan URL; penghapusan tanda baca *punctuation*; *casefolding*; penghapusan *stopwords*, proses tokenisasi, dan proses *POS Tagging*. *POS Tagging* adalah penjelasan kata-kata dengan label yang sesuai berdasarkan konteks kemunculannya. Informasi tentang *Part-Of-Speech* dalam dokumen perlu diekstraksi sehingga skor *SentiwordNet* dapat diterapkan secara akurat.

2. *Sentiwordnet Interpretation Phase*

Untuk mengidentifikasi seberapa positif dan negatif dari kata tertentu dalam teks, setiap kata dalam leksikon *sentiwordnet* skor positif dan negatif dihitung dengan mendapatkan rata-rata untuk entri menurut kategori (Kata Sifat, kata Keterangan, Kata Benda, dan Kata Kerja).

Contohnya adalah kata “*Good*” memiliki 4 arti dari kata benda “*Good*”, 21 arti kata sifat “*Good*”, oleh karena itu akan memiliki tiga nilai positif, dan negatif [9].

3. Menghitung *Term Score Summation*

Untuk menghitung *term score summation* dihitung dengan menjumlahkan skor positif dan negatif untuk setiap istilah yang ditemukan dalam *review*, dihitung untuk mendapatkan skor positif dan negatif untuk semua kata *review*. Kemudian ditentukan sentimen *review* berdasarkan skor mana yang memiliki nilai tertinggi sebagai nilai input *term score summation* jika perbandingan kedua nilai menghasilkan nilai positif lebih tinggi maka skor kata akan ditambahkan sebesar 1 dan dibagi dengan 2, namun jika hasil perbandingan negatif lebih tinggi maka skor negatif tersebut yang akan menjadi nilai input *term score summation*.

4. Menghitung Nilai Kalimat

Skor kalimat dihitung berdasarkan skor kata yang membentuk kalimat dan indeks panjang yang terkait dengan setiap kalimat sesuai persamaan 1. Semakin pendek kalimatnya, semakin mudah untuk menentukan artinya. Jadi, untuk kalimat dengan kurang dari 10 kata, kami mengaitkan indeks panjang 1,4, untuk kalimat yang memiliki 10 hingga 15 kata, kami mengaitkan indeks panjang 1,2, dan untuk kalimat dengan kurang dari 20 kata kami mengaitkan indeks panjang 1,1. Jika tidak, indeks panjangnya adalah 1.

$$\text{nilai_kalimat} = \frac{\sum_{i=0}^n \text{term_score_summation}(i)}{\text{jumlah_kata}} * \text{ukuran indeks} \dots (1)$$

Dimana :

term_score_summation(i) = total nilai *term*-(*i*) pada satu kalimat

jumlah_kata = total jumlah kata pada satu kalimat

ukuran indeks = ukuran nilai indeks yang di hitung berdasar jumlah kata pada satu kalimat

5. Menghitung Nilai Dokumen

Nilai atau skor dokumen digunakan untuk mendapatkan hasil akhir polaritas teks dihitung skor rata-rata kalimat menggunakan persamaan 2.

$$\text{Nilai Dokumen} = \frac{\sum_{i=0}^n \text{Nilai Kalimat}(i)}{\text{Jumlah Kalimat}} \dots (2)$$

Dimana :

Nilai Kalimat(i) = banyaknya nilai kalimat sebanyak-(*i*) dalam satu kalimat *review*

Jumlah Kalimat = banyaknya jumlah kalimat pada setiap dokumen *review*

6. Menentukan Hasil Rating Sistem

Pada tahap ini, hasil rating sistem dikonversi menjadi rating sesuai ketentuan rating 1 memiliki nilai 0-0,2; rating 2 memiliki nilai 0,21-0,4; rating 3 memiliki nilai 0,41-0,6; rating 4 memiliki nilai 0,61-0,8; dan rating 5 memiliki nilai 0,81-1.

7. Menentukan Hasil Analisa Setimen

Tahap Hasil Analisa Sentimen digunakan untuk mengevaluasi hasil rating sistem dari setiap dokumen didasarkan dari rating dokumen tersebut pada *amazon*. Pengujian sistem dengan menggunakan *Confussion Matrix*.

HASIL DAN PEMBAHASAN

Hasil rating dan klasifikasi sentimen dari 5 dokumen ditunjukkan pada Tabel.1. Sedangkan hasil pengujian sistem dari 5 dokumen dengan menggunakan *confussion matrix* ditunjukkan pada Tabel 2. Berdasarkan hasil pengujian sistem, rata-rat akurasi dari sistem analisa sentimen untuk produk HP Samsung pada situs *amazon* adalah 51,33 %.

Tabel 1. Hasil Rating dan Klasifikasi Sentimen Tiap Dokumen

Dokumen	Nilai Dokumen	Rating Uji	Keterangan
1	0.175	1	Negatif
2	0.726	4	Positif
3	0.468	2	Negatif
4	0.141	1	Negatif
5	0.455	2	Negatif

Tabel 2. Hasil Pengujian Sistem Tiap Dokumen

Dokumen ke -	Judul	Pengujian Benar	Pengujian Salah	Akurasi
1	SAMSUNG Galaxy S21 Ultra 5G Factory Unlocked Android Cell Phone 128GB US Version Smartphone Pro-Grade Camera 8K Video 108MP High Res, Phantom Silver	18	12	60%
2	Samsung Electronics Samsung Galaxy S21 5G Factory Unlocked Android Cell Phone US Version 5G Smartphone Pro-Grade Camera, 8K Video, 64MP High Res 128GB, Phantom Gray (SM-G991UZAAXAA	16	14	53,33 %
3	SAMSUNG Galaxy S21 Ultra 5G Factory Unlocked Android Cell Phone 256GB US Version Smartphone Pro-Grade Camera 8K Video 108MP High Res, Phantom Black	17	13	56,66 %
4	Samsung Galaxy S21 Ultra 5G, 128GB Black - Unlocked	14	16	46,66 %
5	Samsung Galaxy S21 5G G991U 128GB GSM/CDMA Unlocked Android Smartphone - Phantom Gray	14	16	46,66%

Kesimpulan

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa sistem analisa sentiment yang dibangun dengan menerapkan lexicon based berdasarkan sentiwordnet dapat

mengklasifikasikan dokumen review produk handphone pada web amazon. Hasil analisa sentiment berdasarkan *review* 5 produk menunjukkan nilai sentimen negative dengan nilai 0,26908. Tingkat akurasi sistem analisa sentimen rata-rata sebesar 51,33%. Berdasarkan nilai akurasi yang masih rendah, pada pengembangan selanjutnya dapat ditambahkan kelas netral pada sistem pengklasifikasian sentimen.

DAFTAR PUSTAKA

- [1] A. Bhatt, A. Patel, H. Chheda, and K. Gawande, "Amazon Review Classification and Sentiment Analysis," *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 6, pp. 5107–5110, 2015.
- [2] A. Rahman and M. S. Hossen, "Sentiment Analysis on Movie Review Data Using Machine Learning Approach," *2019 Int. Conf. Bangla Speech Lang. Process. ICBSLP 2019*, pp. 27–28, 2019.
- [3] B. Saberi and S. Saad, "Sentiment analysis or opinion mining: A review," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 5, pp. 1660–1666, 2017.
- [4] A. Cernian, V. Sgarciu, and B. Martin, "Sentiment analysis from product reviews using SentiWordNet as lexical resource," *Proc. 2015 7th Int. Conf. Electron. Comput. Artif. Intell. ECAI 2015*, pp. 15–18, 2015.
- [5] R. Feldman and J. Sanger, *The Text Mining Handbook*. 2006.
- [6] B. Ohana and B. Tierney, "Sentiment classification of reviews using SentiWordNet," *9th. IT T Conf.*, no. January 2009, 2009.
- [7] C. C. Aggarwal and C. X. Zhai, "Mining text data," *Min. Text Data*, vol. 9781461432234, pp. 1–522, 2013.
- [8] S. Wei and C. P. Soon, "Genetic algorithm-based text clustering technique," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4221 LNCS, pp. 779–782, 2006.
- [9] A. Hamouda and M. Rohaim, "Reviews classification using sentiwordnet lexicon," *World Congr. Comput. Sci. ...*, vol. 2, no. 2, pp. 120–123, 2011.