

SENTIMENT ANALYSIS REVIEW NOVEL “GOODREADS” BERBAHASA INDONESIA MENGGUNAKAN NAÏVE BAYES CLASSIFIER

Diana Ikasari¹, Widiastuti²

^{1,2}Universitas Gunadarma
Margonda Raya 100 Depok

d_ikasari@staff.gunadarma.ac.id, widiastuti@staff.gunadarma.ac.id

ABSTRAK

Novel *best seller* atau novel yang terlaris adalah novel yang banyak diminati oleh pembaca. Orang tidak bisa melihat sebuah novel didefinisikan bagus hanya berdasarkan judul, karena terkadang judul tidak banyak bercerita tentang isi. Novel akan menjadi *best seller* setelah orang banyak membaca kemudian memberikan *review*/testimoni terhadap novel tersebut. *Review* dari pembaca akan memberikan peningkatan terhadap minat untuk membaca novel tersebut apabila *review* bernilai bagus. *Review* tentang sebuah produk sangat penting dalam menentukan pembelian produk, analisis sentimen merupakan proses memahami, mengekstrak dan mengelola data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Hasil dari analisis sentimen adalah bentuk penilaian berkecenderungan positif atau negatif dari *review* pengguna yang dapat memberikan informasi guna pengambilan keputusan untuk membaca atau membeli novel tersebut. Data *review* novel diambil melalui situs *goodreads.com* yang memiliki ribuan *review* tentang sebuah novel. Analisis sentimen dalam penelitian ini menggunakan *naïve bayes classifier* sebagai metode klasifikasi.

Kata Kunci: Goodreads, Novel, *Review*, Analisis Sentimen

ABSTRACT

Best seller novels are novels that in great demand by readers. Someone cannot see that a novel is defined as good only by the title, because sometimes the title doesn't tell much about the content. The novel will become a best seller after people read a lot and then provide reviews / testimonials on the novel. Reviews from readers will provide an increase in interest in reading the novel if the review is good. A review of a product is very important in determining product purchases, sentiment analysis is the process of understanding, extracting and managing textual data automatically to obtain sentiment information contained in an opinion sentence. The result of sentiment analysis is a form of assessment with positive or negative tendencies from user reviews which can provide information for decision making to read or buy the novel. Novel review data is taken through the *goodreads.com* site which has thousands of reviews about a novel. Sentiment analysis in this study uses the naïve Bayes classifier as a classification

Keyword: Goodreads, Novel, *Review*, Sentiment

PENDAHULUAN

Novel merupakan jenis buku yang paling digemari hampir semua kalangan usia, terbukti dengan adanya beragam jenis *genre* novel yang ditentukan oleh klasifikasi umur. Ada novel anak-anak, remaja (*teen-lit* n *chick-lit*), serta *genre* dewasa dan umum. Kepopuleran sebuah novel dapat diketahui dengan melihat daftar buku *best seller* di sebuah toko buku. Kebanyakan daftar tersebut dipenuhi oleh berbagai jenis novel. Novel *best seller* atau novel yang terlaris ini adalah novel yang banyak diminati oleh pembaca. Orang tidak bisa melihat sebuah novel atau buku didefinisikan bagus hanya berdasarkan judul

buku karena terkadang judul buku tidak banyak bercerita tentang isi buku. Umumnya novel akan menjadi *best seller* setelah orang banyak membaca kemudian memberikan testimoni terhadap novel tersebut. Sebelum konsumen memutuskan untuk memilih, sebaiknya konsumen atau pengguna mengetahui detail tentang isi dari buku yang akan dibaca. Hal ini dapat dipelajari berdasarkan testimoni dan opini atau hasil *review* dari pengguna yang pernah membaca buku tersebut. Testimoni akan memberikan peningkatan terhadap minat untuk membaca buku tersebut apabila hasil testimoni bernilai baik atau merekomendasikan bahwa buku tersebut baik dan bagus untuk dibaca.

Saat ini konsumen/pengguna yang menulis tentang testimoni dan opini atau *review* tentang pengalaman dalam membaca suatu buku secara *online* semakin meningkat.

Berdasarkan survei terhadap lebih dari 2000 penduduk dewasa di Amerika, diperoleh data bahwa 81% dari pengguna internet atau sekitar 60% dari penduduk Amerika, sudah melakukan pencarian produk secara *online* minimal satu kali, di antara pembaca *review online* tentang restoran, hotel dan jasa (misalnya agen perjalanan atau dokter) sebanyak 87% melaporkan bahwa ulasan atau *review* atas suatu produk memiliki pengaruh yang tinggi dalam memutuskan apa yang akan dibeli. Sebuah sistem dibutuhkan untuk dapat mengerti akan opini tersebut dengan semakin banyaknya kegiatan *opinion mining* dan analisis sentimen (B. Pang dan L. Lee, 2008).

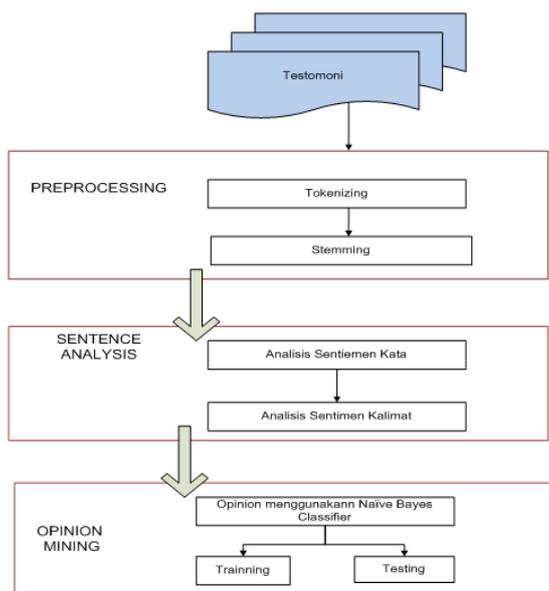
Testimoni mengenai sebuah produk sangat penting dalam menentukan pembelian sebuah produk. Membaca *review* atau sebuah opini secara keseluruhan dapat memakan waktu lama, namun jika hanya sedikit *review* yang dibaca maka evaluasi akan menjadi bias. Klasifikasi sentimen bertujuan mengatasi masalah ini secara otomatis mengelompokkan *review/testimoni* pengguna menjadi opini positif atau negatif (Z. Zhang, Qiang Ye, Yijun Li, 2011). Kalimat yang terkandung dalam testimoni dapat berupa kalimat yang berpandangan negatif atau positif. Testimoni merupakan *opinion mining* atau termasuk dalam analisis sentimen, yaitu merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Pada penelitian ini analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap suatu produk karya sastra dalam hal ini yaitu novel, apakah cenderung berpandangan atau beropini negatif atau positif agar pada akhirnya konsumen/pengguna dapat mengetahui tanggapan atau testimoni tentang novel tersebut secara cepat dan tepat. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat, seperti halnya di Amerika banyak terdapat perusahaan yang fokus pada layanan analisis sentimen (Liu, B, 2010).

Beberapa situs untuk pencarian novel *online* di antaranya adalah goodreads.com, belbuk.com, bukukita.com, ebook.bike dan ibookile.com. Penelitian ini memilih situs goodreads.com sebagai data penelitian karena pada Goodreads dapat terlihat ribuan *review* tentang sebuah judul novel, sehingga dapat menjadi bahan pertimbangan dalam memilih novel yang akan dibaca. Goodreads merupakan salah satu situs buku *online* yang menyediakan fasilitas berbagi rekomendasi buku antar sesama anggota, mencari informasi tertentu tentang buku maupun penulis buku serta dapat menghubungkan orang-orang yang memiliki ketertarikan pada buku yang sama.

Naïve Bayes Classifier merupakan salah satu algoritma klasifikasi yang memiliki performa cukup baik ketika digunakan untuk melakukan klasifikasi teks. Penelitian Lyu Kigon, Kim Hyeoncheol (2016) menunjukkan bahwa nilai *precision* klasifikasi *spam* yang dilakukan oleh *web service* yang menggunakan *Naïve Bayes Classifier* adalah 93.3%. Pada penelitian lain Zhang, L., Hua, K., Wang, H., Qian, G., & Zhang, L (2013) melakukan analisis sentimen terhadap dokumen teks yang menghasilkan akurasi sebesar 83%. Pada penelitian ini digunakan metode *Naïve Bayes* untuk menentukan makna kalimat berdasarkan *opinion mining*. Pada klasifikasi sentimen, dilakukan pendefinisian data *training* yang digunakan untuk melatih sistem dengan kemampuan belajar dari data yang dapat mengajarkan dirinya sendiri untuk tumbuh dan berubah bila diberikan data baru. Hasil dari penelitian ini adalah klasifikasi makna positif dan negatif dari *review* terhadap novel berdasarkan sentimen analisis menggunakan metode *Naïve Bayes*.

METODE PENELITIAN

Metode penelitian yang digunakan pada penelitian menggambarkan proses pengolahan sentimen analisis, yaitu mengolah data testimoni dimulai dari *preprocessing* sampai ke tahap klasifikasi opini berkecenderungan positif atau negatif, tahapan penelitian terlihat pada gambar 1.



Gambar 1. Tahapan Penelitian Pengolahan Analisis Sentimen Novel

Gambar 1 merupakan tahapan yang dilakukan dalam penelitian ini yang dimulai dari melakukan pemrosesan dari data testimoni buku yang sudah ada sebagai tahap *preprocessing* untuk menghasilkan kata dasar melalui *tokenizing* dan *stemming*. Pada tahap kedua adalah *Sentence Analysis* dengan melakukan proses analisis sentimen kata dan proses analisis sentimen kalimat untuk melakukan penilaian kata-kata yang sudah dilakukan pada proses *stemming* dalam *preprocessing*. Dilanjutkan dengan tahap yang ketiga yaitu proses penentuan makna kalimat berdasarkan *opinion mining* dengan menggunakan metode Naïve Bayes. Pada pengklasifikasian sentimen dilakukan pendefinisian data *training* yang digunakan untuk melatih sistem dengan kemampuan belajar dari data, mengembangkan program komputer yang dapat mengajarkan dirinya sendiri untuk tumbuh dan berubah bila diberikan data baru. Setelah pembuatan data *training* dilanjutkan dengan menguji data *training* untuk melihat keberagaman sehingga bisa diterapkan di dalam korpus kata apa saja yang ditemukan. Bila sudah *fixed* terlihat, dilanjutkan dengan pengujian data sebagai bentuk pembuktian dari langkah yang sudah dilakukan dengan melakukan data *testing*.

Pengumpulan Data Set Testimoni

Pengumpulan sumber informasi tentang novel yang menjadi bagian dari data-data yang diperlukan dari penelitian ini dengan

mengumpulkan uraian *review* dari pengguna terhadap novel melalui situs *goodreads.com*.

Preprocessing

Tahapan *preprocessing* dilakukan berdasarkan beberapa langkah sebagai berikut:

1. Pemilihan ulasan yang digunakan yaitu ulasan yang memiliki makna positif dan negatif, kemudian teks ulasan dimasukkan, maka akan dilakukan penghapusan *delimiter* yaitu karakter angka dan karakter simbol kecuali karakter huruf, kemudian merubah semua karakter huruf menjadi huruf kecil melalui proses *to LowerCase*, proses ini disebut *Tokenizing*.
2. Kemudian dilakukan proses *stopword removal*. Setiap kata-kata yang ada di dalam ulasan tersebut akan dibandingkan dengan daftar *stopword* yang ada di dalam *database*. Apabila ditemukan kata yang sama, maka kata tersebut akan dihapus dari pernyataan.
3. Semua kata yang ada di dalam ulasan dibandingkan dengan daftar kata dasar yang ada di dalam *database*. Apabila kata tidak ditemukan, maka sistem akan melakukan proses *stemming Enhanced Confix Stripping* pada kata tersebut.
4. Proses *text pre-processing* selesai.

Sentence Analysis

Proses pada tahap *Sentence Analysis*, terdiri atas dua prosedur yaitu prosedur analisis sentimen kata dan prosedur analisis sentimen kalimat, prosedur sentimen kata dilakukan untuk mendefinisikan setiap kata yang merupakan hasil dari proses *stemming* yang sudah dilakukan pada *preprocessing* untuk didefinisikan apakah terdapat pada *database* kata cenderung +/-, serta melakukan pencocokan kata terhadap *database* sinonim kata yang dimiliki, apabila kata tersebut tidak ada yang memiliki kecocokan dan sudah berada pada data yang terakhir maka kata tersebut akan disimpan di dalam *database* kata cenderung +/- sebagai kata baru yang menambah koleksi kata yang bermakna +/-.

Berdasarkan hasil dari pendefinisian kata yang memiliki kecenderungan positif atau negatif yang telah dihasilkan dari proses analisis sentimen kata dengan tersimpannya kata dalam *database temporary*, proses dilanjutkan pada prosedur analisis sentimen kalimat, yaitu melakukan penentuan makna dari kalimat yang terbentuk dari kata yang bernilai cenderung

positif atau negatif yang mendefinisikan apakah kalimat pembentuk testimoni bernilai positif atau negatif dengan melakukan perhitungan jumlah kata yang bernilai positif dan yang bernilai negatif.

Opinion Mining Menggunakan Naïve Bayes Classifier

Algoritma *Naive Bayes Classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat.

Algoritma Pelatihan

1. Jumlah semua *token* ← jumlah semua kata yang unik dari dokumen
2. Untuk setiap kelas sentimen lakukan:
 - a. Jumlah *record* pada kelas *j* ← jumlah *record* yang berada pada kelas *j*
 - b. Hitung $P(\text{sentiment}_j)$ dengan persamaan (1)

$$P(V_j) = \frac{|doc_j|}{Contoh}$$

- c. Untuk setiap kata w_k pada daftar semua token lakukan hitung $P(\text{kata}_k|\text{sentiment}_j)$ dengan persamaan (2)

$$P(W_k|V_j) = \frac{n_k + 1}{n + |vocabulary|}$$

Algoritma Klasifikasi Naïve Bayes Classifier

1. Input pesan (dokumen) yang akan diketahui sentimen (klasifikasi)-nya.
2. Hasilkan probabilitas untuk masing-masing kelas sesuai dengan persamaan (1) dengan menggunakan $P(\text{sentiment}_j)$ dan $P(\text{kata}_k|\text{sentiment}_j)$ yang telah diperoleh dari pelatihan.
3. Probabilitas kelas maksimum adalah kelas sentimen terpilih hasil klasifikasi.

Proses inti adalah proses klasifikasi untuk menentukan sebuah kalimat sebagai anggota opini positif atau sebagai anggota opini negatif berdasarkan nilai perhitungan probabilitas *Bayes* yang lebih besar. Jika hasil probabilitas *Bayes* kalimat tersebut untuk opini positif lebih besar maka kalimat tersebut masuk kategori opini positif demikian juga sebaliknya. Mengukur kinerja algoritma digunakan rumus akurasi klasifikasi sebagai berikut:

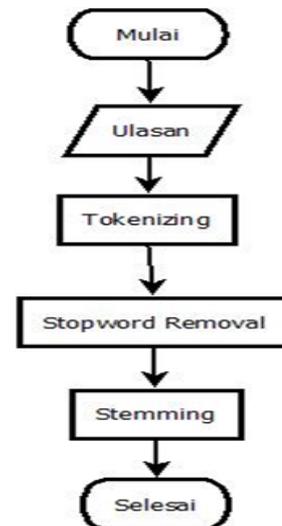
$$Akurasi = \frac{Jumlah\ klasifikasi\ benar}{Jumlah\ dokumen\ diklasifikasi} \times 100\%$$

HASIL DAN PEMBAHASAN

Skenario ujicoba dari algoritma yang telah dijelaskan pada metode penelitian dilakukan dengan tahapan:

1. *Preprocessing* teks.
2. Perhitungan sentimen kata dan sentimen kalimat testimoni.
3. Menampilkan hasil perhitunga testimoni.

Preprocessing teks: diawali dengan melakukan input terhadap data testimoni dalam bentuk kalimat, kemudian dilaukan proses *tokenizing*, *stopword removal* dan dilanjutkan dengan proses *stemming*, proses terlihat pada gambar 2.



Gambar 2. Flowchart Preprocessing Teks

Contoh proses *preprocessing* teks:

- a. Inputan kalimat Testimoni sebagai berikut :

Buku Ini Membantu Kita Memahami Bahwa Kita Berharga.

- b. *Tokenizing*: proses perubahan menjadi huruf kecil, dan kalimat inputan berubah menjadi huruf kecil seluruhnya.

buku ini membantu kita memahami bahwa kita berharga.

- c. Proses penghilangan semua tanda baca, menghasilkan pernyataan tanpa adanya tanda baca, proses *tokenizing* selesai, hasil sebagai berikut:

buku ini membantu kita memahami bahwa kita berharga

- d. Proses *Stopword Removal* menggunakan *output* dari *tokenizing*, hasil sebagai berikut:

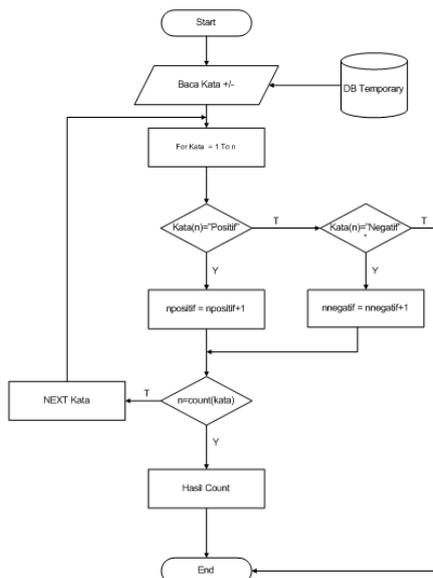
buku membantu memahami berharga

e. Proses *stemming* menghasilkan kata dasar/*root word* yang disimpan dalam *database root word*

buku bantu paham harga

Perhitungan sentimen kata dan sentimen kalimat testimoni.

Pada proses ini dilakukan penentuan makna kalimat yang terbentuk dari kata yang cenderung positif atau negatif, yang akhirnya mendefinisikan apakah kalimat pembentuk testimoni bernilai positif atau negatif, terlihat pada gambar 3.



Gambar 2. Flowchart analisis sentiment kalimat

Hasil pengujian data terlihat pada tabel 1. Nilai yang diproses pada program penilaian sentimen adalah: n_{pos} & n_{neg} , kedua nilai tersebut digunakan untuk penilaian komentar dengan menggunakan rumus total nilai dengan jumlah sebagai banyak kata dalam komentar setelah proses *stemming*, adalah:

$$nilai_{total} = \frac{n_{positif} + (-n_{negatif})}{Jumlah}$$

Variabel nilai_{total} akan dibuat sebagai acuan untuk klasifikasi komentar, “apakah bernilai positif atau negatif”, dengan logika:

- Jika n_{pos} & n_{neg} bernilai 0 maka komentar adalah NETRAL
- Jika nilai_{total} >0, maka komentar bernilai positif
- Jika nilai_{total} <0, maka komentar bernilai negatif

Implementasi klasifikasi komentar ini menggunakan percabangan *else*, dan

selanjutnya nilai komentar tersebut digunakan untuk mengurutkan buku.

Contoh: Hasil *stemming* kalimat testimoni adalah sebagai berikut:

buku janji baik nama nama karakter afrika sulit lacak siapa lupa nama karakter sulit nama afrika

Jumlah kata = 16

$n_{positif}$ = baik (1)

$n_{negatif}$ = sulit, lupa, sulit (-3)

$$nilai_{total} = \frac{1 + (-3)}{16} = -\frac{2}{16} = -0,125$$

Jika nilai total < 0, maka nilai sentimen adalah negatif.

Tabel 1. Skenario Hasil Uji Coba Analisis Sentimen

No	Judul Buku	Kalimat Testimoni	Hasil Stemming	Nilai dan Score Sentimen
1.	Things Fall Apart V	Buku ini membuat saya tidak menghormati budaya yang saya tahu sedikit sebelum membacanya.	buku tidak hormat budaya baca	NEGATIF -0.2
2.	Komik 100 Kebiasaan Nabi	Lucu dan mendidik. Hadiah yang layak diberikan untuk anak.	lucu didik hadiah layak anak	POSITIF 0.4
3.	Komik 100 Kebiasaan Nabi	Mungkin inilah komik Islami yang pertama yang betul-betul lahir dari khazanah kultur Indonesia dan mengangkat kearifan lokal Bangsa Indonesia.	komik islami lahir khazanah kultur indonesia angkat arif lokal bangsa indonesia	POSITIF 0.1818
4	Things Fall Apart V	Nah, ini mulai terlihat seperti sebuah buku yang menjanjikan	buku janji baik nama nama karakter afrika sulit lacak siapa lupa	NEGATIF -0.125

		an baik, tapi nama-nama karakter adalah Afrika dan itu sangat sulit untuk melacak siapanya, Anda lupa karena nama karakter yang begitu sulit untuk diingat, menggunakan nama Afrika "	nama karakter sulit nama afrika	
--	--	---	---------------------------------	--

SIMPULAN DAN SARAN

Pada penelitian ini telah dilakukan perhitungan nilai analisis sentimen terhadap *review* novel berbahasa Indonesia yang terdapat pada situs goodreads.com menggunakan metode Naïve Bayes Classifier. Hasil analisis sentimen ini memberikan nilai *review* berkecenderungan positif atau *review* berkecenderungan negatif. Hasil perhitungan *review* tersebut nantinya dapat digunakan untuk melakukan perbandingan novel sehingga pengguna akan lebih mudah menentukan novel yang baik

untuk dibaca atau dibeli berdasarkan rekomendasi *review* dari pengguna lain. Penelitian ini hanya terbatas pada tahap perhitungan nilai analisis sentimen kalimat dan dapat dilanjutkan nantinya menjadi perbandingan atau rekomendasi novel yang lebih baik.

DAFTAR PUSTAKA

- Liu, B. (2010). Hand Book of natural Language Processing, Chapter Sentiment Analysis and Analysis, 2nd Edition, Chapman & Hall/CRC Press.
- Lyu Kigon, Kim Hyeoncheol. (2016). Sentiment Analysis Using Wrd Polarity of Social Media. Springer Science, Businee Media, New York.
- Mohammad Ali Abbasi, Huan Liu. (2013). Measuring User Credibility in Social Media. Proceeding 6th International Conference, SBP, Washington DC, USA.
- Pang, B., Lee, L., & Vithyanathan, S. (2008). Sentiment Classification Using Machine Learning Techniques.
- Zhang, L., Hua, K., Wang, H., Qian, G., & Zhang, L. (2014). Sentiment Analysis on Reviews of Mobile Users. Procedia Computer Science, pp 458-465.
- Ziqiong Zhang, Qiang Ye, Zili Zhang, Yijun Li. (2011). Sentiment Classification of Internet Restaurant Reviews Written in Cantonese. Expert System with Application.