

Pengaruh Proses *Stemming* Pada Kinerja Analisa Sentimen Pada *Review* Buku

Luvia Friska Narulita

Program Studi Teknik Informatika, Universitas 17 Agustus 1945 Surabaya

Email : luvianarulita@gmail.com

Abstrak

Analisa sentimen dapat dilakukan pada dokumen teks yang tidak terstruktur termasuk pada dokumen tinjauan buku. Tahapan pra pemrosesan menjadi tahapan yang penting untuk dilakukan sebelum analisa sentimen dilakukan. Tahapan pra pemrosesan yang meliputi proses *stemming* menjadi penting dikarenakan proses tersebut dapat meningkatkan ketepatan analisa sentimen pada dokumen teks.

Kata Kunci: *stemming, analisa sentimen, k nearest neighbor, review buku.*

Pendahuluan

Internet saat ini banyak digunakan untuk pencarian informasi. Banyak sekali informasi yang disediakan di berbagai situs web, baik informasi mengenai politik, informasi mengenai produk maupun situs web yang menyediakan tempat untuk penjualan berbagai produk.

Informasi mengenai produk, film atau buku dapat dijumpai di situs – situs yang menjual produk atau tiket pertunjukan film tersebut. Informasi yang didapatkan diantaranya adalah nama, harga dan penjelasan singkat mengenai produk tersebut. Saat ini banyak situs web yang menyertakan review atau tinjauan dari pengguna produk yang ditampilkan pada situs tersebut. Hal tersebut dikarenakan menurut penelitian dari (Maslowska, Malthouse, & Viswanathan, 2017) menunjukkan bahwa adanya tinjauan terhadap barang dan buku mempengaruhi keputusan seseorang untuk melakukan pembelian barang.

Tinjauan yang ditulis oleh pengguna di internet, termasuk tinjauan yang ditulis oleh pembeli produk, penonton film atau pembaca buku termasuk ke dalam data tidak terstruktur. Pengolahan data tidak terstruktur untuk memperoleh informasi dari susunan data tersebut dapat dilakukan dengan menggunakan teknologi teks mining.

Pada pengolahan informasi dengan teks mining diperlukan tahapan *preprocessing* yang berfungsi untuk

mempersiapkan data sebelum data tersebut siap untuk diolah. (Vijayarani, Ilamathi, & Nithya) telah melakukan penelitian dengan membagi tahap *preprocessing* menjadi tiga tahapan kunci, yaitu penghapusan *stop word*, *stemming* dan algoritma TF/IDF. Setiap tahapan dalam tahap *preprocessing* memiliki pengaruh masing – masing pada hasil pengolahan data. Pengaruh tahapan tersebut menjadi dasar peneliti untuk melakukan penelitian mengenai pengaruh tahapan *stemming* pada analisa sentimen pada tinjauan buku.

Data yang digunakan pada penelitian ini adalah data tinjauan buku yang didapatkan melalui situs goodreads.com untuk buku berbahasa Indonesia. Kategori buku yang digunakan sebagai data penelitian adalah novel remaja. Teks tinjauan buku yang diambil untuk pengolahan data adalah teks berbahasa Indonesia sejumlah 72 data dengan 36 data tinjauan positif dan 36 data tinjauan negatif.

Kajian Literatur

Stemming digunakan untuk menghilangkan imbuhan – imbuhan yang ada dalam kata – kata sehingga didapatkan kata dasar. Algoritma yang digunakan untuk proses *stemming* dapat berbeda jika digunakan untuk bahasa yang berbeda. Diantara algoritma yang telah dikembangkan untuk proses *stemming* dokumen berbahasa Indonesia adalah algoritma Nazief & Adriani dan Algoritma Porter (Wahyudi, Susyanto, & Nugroho).

Algoritma Nazief & Adriani dikembangkan pada tahun 1996 dan dipublikasikan secara internal pada fakultas Teknik Komputer, Universitas Indonesia. Algoritma Nazief & Adriani ditulis oleh B.A.A Nazief dan M. Adriani. Jelita Asian B.Comp. Sc. (Hons) menulis tentang tehnik – tehnik yang efektif untuk *text retrieval* dalam Bahasa Indonesia. Dalam dokumen tesis tersebut juga dituliskan mengenai langkah – langkah algoritma *stemming* Nazief & Adriani (Asian, 2007) yang meliputi:

- a. Langkah awal yang dilakukan adalah dengan menyiapkan daftar kata – kata dasar dalam Bahasa Indonesia yang bisa disimpan dalam format kamus kata dasar Bahasa Indonesia.
- b. Setelah kamus kata dasar Bahasa Indonesia tersedia, maka pada langkah selanjutnya yang dilakukan oleh algoritma ini adalah membandingkan kata yang diinputkan dengan setiap kata yang ada pada kamus kata dasar, jika kata yang diinputkan terdapat pada kamus, maka kata yang diinputkan tersebut dianggap sebagai kata dasar dan proses pencarian berhenti.
- c. Langkah berikutnya adalah dengan menghapus *inflectional suffixes* atau imbuhan yang tidak mempengaruhi cara pembacaan kata dasar yang diberikan imbuhan tersebut. *Inflectional suffixes* terdiri dari dua macam, yaitu *inflectional particle suffixes* dan *possessive pronoun suffixes*. Contoh imbuhan partikel atau *particle suffixes* tersebut adalah “-lah”, “-kah”, “-tah” atau “-pun”. Sedangkan contoh penggunaan imbuhan tersebut adalah “makanlah”. Kata dasar “makan” tidak berubah cara pembacaannya ketika diberikan imbuhan “-lah”. Contoh imbuhan *possessive pronoun suffixes* atau imbuhan yang menyatakan kepemilikan adalah “-ku”, “-mu” atau “-nya”. Kedua imbuhan tersebut harus dihilangkan untuk mendapatkan kata dasar. Setelah proses penghilangan imbuhan dan

didapatkan kata dasar, maka langkah selanjutnya adalah membandingkan kata dasar tersebut. Jika kata dasar tersebut terdapat pada kamus kata dasar, maka algoritma berhenti.

- d. Langkah berikutnya adalah menghilangkan imbuhan derivatif seperti “-i”, “-kan”, dan “-an”. Imbuhan tersebut biasanya diletakkan pada akhir kata dan membentuk kata kerja dari kata dasar yang telah diberi imbuhan. Contoh “pulangkan” memiliki kata dasar “pulang” dengan imbuhan “-kan”. Setelah dihilangkan imbuhan yang menyertai kata tersebut, maka langkah selanjutnya adalah membandingkan kata yang didapatkan dengan kamus kata dasar yang ada. Jika kata tersebut ditemukan, maka algoritma berhenti. Jika kata tersebut tidak ditemukan, maka algoritma melangkah ke langkah selanjutnya.
- e. Menghilangkan awalan merupakan langkah berikutnya pada algoritma ini. Awalan yang dihilangkan adalah “be-”, “di-”, “ke-”, “me-”, “pe-”, “se-”, dan “te-”. Langkah ini dihentikan jika:
 - a. Ditemukan pasangan awalan dan akhiran yang telah ditemukan sebelumnya merupakan pasangan yang tidak diperbolehkan. Seperti pasangan awalan “ber” dan akhiran “i” pada sebuah kata.
 - b. Awalan yang ditemukan sama persis dengan awalan yang telah ditemukan sebelumnya
 - c. Tiga awalan telah dihilangkan.
- f. Langkah berikutnya adalah dengan mengidentifikasi tipe awalan dan menghilangkan ambiguitas.
- g. Proses penghilangan awalan merupakan proses berulang, jika sampai proses penghilangan awalan berakhir tetapi kata dasar tidak ditemukan, maka algoritma berhenti.
- h. Jika semua langkah telah dilakukan namun tidak berhasil, maka kata dikembalikan seperti pada awal sebelum kata tersebut dilakukan proses *stemming*.

Analisa Sentimen atau disebut juga sebagai *opinion mining* merupakan bidang yang mempelajari dan menganalisa opini publik, sentimen, evaluasi, penilaian, sikap dan emosi terhadap suatu produk, pelayanan, organisasi, individu, isu, kejadian, topik dan atribut – atribut yang menyertai (Liu, 2012)

Analisa sentimen pada dokumen teks merupakan proses untuk menganalisa kecenderungan suatu teks terhadap sentimen positif atau negatif. Contoh sentimen positif pada suatu kata adalah kata “bagus”, “baik” atau “cantik”. Sedangkan contoh sentimen negatif adalah “jelek” atau “buruk”.

Metode

Terdapat empat cara untuk mengukur Metode yang digunakan pada penelitian ini adalah sebagai berikut:

- a. Pengumpulan data
Data yang digunakan diambil dari situs goodreads.com secara acak dengan mengambil data tinjauan pada novel berbahasa Indonesia dan tinjauan dengan bahasa Indonesia. Pengguna atau pemberi tinjauan pada situs tersebut dapat memberikan nilai berupa bintang dengan rentang nilai satu bintang sampai dengan lima bintang.
- b. Tahap *Preprocessing*
Tahap *preprocessing* terdiri dari tahap penghapusan *stopword* dan tahap *stemming*. Untuk membandingkan hasil analisa sentimen dilakukan perbedaan pada tahap *preprocessing*, yaitu dengan menyertakan tahap *stemming* dan dengan tidak menyertakan tahap *stemming*. Algoritma yang digunakan pada proses *stemming* adalah algoritma Nazief & Ardani. Langkah – langkah algoritma tersebut telah dijelaskan pada bab Tinjauan Pustaka.
- c. Pembobotan istilah (*term*)
Pembobotan istilah digunakan untuk memberikan nilai pada setiap istilah atau kata yang telah dipisahkan.
- d. Penghitungan tingkat kemiripan.

Penghitungan tingkat kemiripan dari setiap kata berfungsi untuk menghitung kemiripan kata dalam kalimat dengan kumpulan kalimat pada data latih. Pada proses ini dihitung tingkat kemiripan data sehingga dapat diketahui tingkat kemiripan kalimat terhadap kumpulan data positif dan data negatif.

- e. Klasifikasi
Proses klasifikasi dilakukan dengan menggunakan aplikasi Weka 3.8 dengan algoritma Naïve Bayes dan K Nearest Algorithm
- f. Pengujian
Pengujian dilakukan dengan membandingkan data uji terhadap data latih dan membandingkan secara langsung hasil yang didapatkan dengan nilai aktual.

Hasil dan Pembahasan

Hasil yang didapatkan untuk proses analisa sentimen dengan menyertakan tahapan *stemming* setelah dilakukan proses klasifikasi dengan menggunakan aplikasi Weka 3.8 ditunjukkan pada grafik berikut



Gambar 4.1: Tingkat Akurasi dengan algoritma KNN dengan k = 1



Gambar 4.2: Tingkat Akurasi dengan algoritma KNN dengan k = 5



Gambar 4.3: Tingkat Akurasi dengan algoritma KNN dengan k = 10



Gambar 4.4: Tingkat *precision* dengan k = 1

Gambar 4.1, 4.2 dan 4.3 menunjukkan tingkat akurasi yang diperoleh. Dalam gambar tersebut ditunjukkan bahwa tingkat akurasi dengan menggunakan proses *stemming* lebih tinggi jika dibandingkan dengan analisa sentimen tanpa proses *stemming*. Dalam gambar tersebut ditunjukkan hasil yang diperoleh dengan *percentage split* 70%, 75% dan 80%.



Gambar 4.5: Tingkat *precision* dengan k = 5

Tabel 4.1.

Tabel Akurasi Tanpa *Stemming*

k	Percentage Split		
	70%	75%	80%
1	45,45	44,44	50,00
5	59,00	50,00	57,14
10	68,18	61,11	64,28

Tabel 4.2.

Tabel Akurasi Dengan *Stemming*

k	Percentage Split		
	70%	75%	80%
1	77,27	77,77	78,57
5	68,18	77,77	71,42
10	77,27	77,77	78,51

Tabel 4.1 dan Tabel 4.2 menunjukkan perbedaan tingkat akurasi yang diperoleh dari analisa sentimen dengan menggunakan proses *stemming* dan tanpa menggunakan proses *stemming*. Dengan menggunakan *classifier* iBK dengan k = 1, k = 5, k = 10 menunjukkan nilai akurasi dengan proses *stemming* lebih tinggi dibandingkan dengan tanpa proses *stemming*.

Gambar 4.4 menunjukkan tingkat *precision* yang diperoleh.



Gambar 4.6: Tingkat *precision* dengan k=10

Tingkat *precision* yang ditunjukkan pada gambar 4.4, gambar 4.5 dan gambar 4.6 adalah tingkat ketepatan sistem dalam melakukan prediksi. Nilai *precision* yang lebih jelas ditunjukkan pada tabel 4.3 dan tabel 4.4.

Tabel 4.3.

Tabel nilai *precision* untuk analisa sentimen tanpa proses *stemming*.

k	Percentage Split		
	70%	75%	80%
1	0,455	0,442	0,576
5	0,618	0,5	0,786
10	0,701	0,781	0,805

Tabel 4.4.

Tabel nilai *precision* untuk analisa sentimen dengan proses *stemming*.

k	Percentage Split		
	70%	75%	80%
1	0,771	0,792	0,857
5	0,751	0,792	0,829
10	0,784	0,846	0,857

Dari tabel 4.3 dan tabel 4.4 terlihat bahwa nilai *precision* untuk proses analisa sentimen dengan menggunakan proses *stemming* lebih tinggi jika dibandingkan dengan analisa sentimen tanpa proses *stemming*.

Kesimpulan

Kesimpulan yang didapatkan dari penelitian yang telah dilakukan adalah bahwa proses *stemming* memberikan peranan penting untuk meningkatkan akurasi analisa sentimen untuk dokumen teks.

Pada penelitian selanjutnya diharapkan dapat dilakukan penelitian dengan menggunakan algoritma *stemming* serta algoritma klasifikasi yang lain sehingga dapat diperoleh perbandingan kinerja di antara beberapa algoritma. Selain itu, data yang digunakan juga dapat diperbanyak sehingga hasil penelitian lebih akurat.

Referensi

- Maslowska, E., Malthouse, E., & Viswanathan, V. (2017). Do Customer Reviews Drive Purchase Decisions? The Moderating Roles of Review Exposure and Price. *Decision Support System*.
- Asian, J. (2007). *Effective Techniques for Indonesian Text Retrieval. Thesis of Doctoral*. Melbourne: RMIT University.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Vijayarani, D., Ilamathi, M., & Nithya, M. (n.d.). Preprocessing Techniques for Text Mining - An Overview. *International Journal of Computer Science & Communication Networks vol 5 no 1*, 7 - 16.

Wahyudi, D., Susyanto, T., & Nugroho, D. (n.d.). Implementasi dan Analisis Algoritma Stemming Nazief & Adriani dan Porter pada Dokumen Berbahasa Indonesia. *Jurnal Ilmiah SINUS*, 49 - 56.