Opinion Mining Pada Twitter Untuk Bahasa Indonesia Dengan metode Support Vector Machine dan Metode Berbasis Lexicon

(Center, Times New Roman 18, maks **12 kata Bhs. Ind. or 10 words in English**  )

**Jan Kristanto\*1, Edi Winarko2**

1,2Institution/affiliation; addres, telp/fax of institution/affiliation

3Jurusan Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta

e-mail: **\***[**1jan\_kristanto@ugm.ac.id**](mailto:1jan_kristanto@ugm.ac.id), [2ewinarko@ugm.ac.id](mailto:2ewinarko@ugm.ac.id)

***Abstrak***

*Dalam proses pengambilan keputusan opini atau pendapat orang lain sering dijadikan acuan. Dengan internet dan sosial media yang ada sekarang banyak orang sering berpendapat,mengeluh atau mengkritik tentang sesuatu hal. Sehingga untuk mengetahui pendapat orang lain tentang sesuatu hal kita tidak perlu bertanya kepada mereka satu per satu. Kita dapat mengembangkan metode yang dapat mencari dan menganalisa teks pada sosial media yang menghasilkan output berupa pendapat tentang sesuatu hal.*

*Metode untuk mendapatkan informasi berupa pendapat yang terdapat dalam suatu teks adalah “Opinion Mining” atau “Sentiment Analysis”. Pada penelitian ini menggunakan metode gabungan Unsupervised yang berbasis pada lexicon berbahasa indonesia dan Supervised yang berbasis pada Support Vector Machine.*

*Hasil yang diperoleh berupa polaritas sentimen apakah negatif , positif, ataupun netral. Dari hasil pengujian pada penelitian ini didapatkan akurasi 77,7%.*

***Kata kunci***— opinion mining, sentiment analysis, lexicon, support vector machine

***Abstract***

*Opinions in the decision-making process are often used as a reference. Nowadays, with the internet and social media there are a lot of people often argue, complain or criticize about something. So as to know what others think of something that we do not need to ask them one by one. We can develop a method that can search and analyze text on social media which produces output in the form an opinion about something.*

*The method to obtain information in the form of opinions contained in the text is an "Opinion Mining" or "Sentiment Analysis". In this research, using a combined Unsupervised method based on Indonesian Lexicon and Supervised based on Support Vector Machine.*

*Results obtained in the form of sentiment is negative polarity, positive, or neutral. From the test results obtained in this study 77,7% accuracy.*

***Keywords***— opinion mining, sentiment analysis, lexicon, support vector machine

1. PENDAHULUAN

O

pini adalah pendapat pribadi yang tidak obyektif dan tidak melalui proses verifikasi (Quirk et al., 1985). Opini orang lain tentang suatu hal menjadi penting dalam pengambilan keputusan. Misalkan seseorang akan membeli sebuah mobil biasanya dia akan bertanya kepada orang lain bagaimana pendapat mereka tentang mobil tersebut. Atau sering juga ketika kita membeli produk tertentu, kita dipersilakan untuk mengisi kuisoner yang berkaitan dengan produk tersebut. Sekarang ini tidak perlu kita langsung bertanya tentang pendapat orang lain secara langsung. Seiring dengan perkembangan internet dan web, telah banyak orang yang berbagi informasi tentang pendapat dan pengalaman bahkan kritik mereka terhadap suatu produk tertentu.

Social media kini telah berkembang dengan pesat, dimulai dari awal Friendster sampai era sekarang ini Facebook, Twitter, Google+ dan Foursquare. Menurut socialbakers.com pengguna Facebook di Indonesia telah mecapai 42.5 juta, yang menjadikan Indonesia menjadi peringkat ke empat pengguna Facebook setelah Amerika, India dan Brazil. Pengguna Twitter di Indonesia juga tidak kalah besar jumlahnya yang mencapai 19.5 juta (Semiocast, 2012) yang menempati peringkat kelima setelah Amerika, Brazil, Japan dan U.K. Perkembangan yang begitu pesat ini membuat orang-orang berkomunikasi dengan mudah. Mereka saling berbagi informasi, baik berbagi tentang aktifitas sehari-hari mereka ataupun memberikan pemberitahuan pada komunitas mereka, bahkan keluhan mereka terhadap suatu hal.

Begitu populernya social media, pengguna social media sekarang ini bukan hanya kalangan pribadi atau perseorangan saja. Tapi kalangan corporate juga memanfaatkan social media sebagai media komunikasinya dengan masyarakat. Seperti pengenalan produk baru, penawaran produk, bahkan para legislative juga mempromosikan diri mereka melalui social media. Selain yang dianggap lebih murah dari segi biaya, juga lebih tepat pada sasaran.

Berdasarkan pada paparan diatas terdapat potensi besar pada social media untuk dilakukan pengekstrakan opini atau yang disebut opinion mining, untuk mengetahui pendapat orang lain tentang suatu hal, yang dapat dimanfaatkan selama proses pengambilan keputusan. Beberapa penelitian yang sudah ada sekarang ini kebanyakan menggunakan machine learning baik dengan naïve bayes (Alliandu dan Winarko, 2012) , (Pang et al., 2002), (Franky dan Manurung, 2008), (Go et al., 2009), (Pak dan Paroubek, 2010) , Support Vector Machine (Pang et al., 2002) , (Franky dan Manurung, 2008), (Zhang et al., 2008), (Go et al., 2009) dan Maximum Entropy (Go et al., 2009) ataupun yang berbasiskan lexicon (Komarsilam dan Winarko, 2012). Pada penelitian ini akan dilakukan opinion mining dengan mengkombinasikan antara pendekatan berbasis machine learning dan berbasiskan lexicon. Metode yang akan digunakan untuk mewakili pendekatan machine learning adalah Support Vector Machine (SVM), yang menurut paper Zhang et al. (2008) sering memberikan akurasi yang terbaik dibanding dengan metode learning yang lainnya. Dengan melakukan kombinasi antara metode Support Vector Machine (SVM) dan berbasiskan lexicon diharapkan dapat meningkatkan akurasi pada proses pengklasifikasian opini.

2. METODE PENELITIAN

Makalah hendaknya memuat tulisan yang berisi 1.**Pendahuluan, 2. Metode Penelitian (bisa meliputi analisa, arsitektur, metode yang dipakai untuk menyelesaikan masalah, implementasi), 3. Hasil dan Pembahasan, 4. Kesimpulan dan 5. Saran** (future works) yg berisi penelitian lanjut di masa mendatang. Pada setiap paragraph bisa terdiri dari beberapa subparagraph yang dituliskan dengan penomoran angka arab seperti yang ditunjukkan section berikut ini. Jumlah halaman **minimum 10 halaman dan maksimum 12 halaman** ukuran **A4.**

## 2.1 Tahapan Review

Harap mengirimkan naskah anda secara elektronik untuk direview sebagai attachments e-mail. Ketika anda mengirimkan dokumen naskah versi awal dalam format *Word*.doc satu kolom, termasuk gambar dan tabel.

## 2. 1.1 Gambar dan tabel

Semua tabel dan gambar yang anda masukkan dalam dokumen harus disesuaikan dengan urutan 1 kolom atau ukuran penuh satu kertas, agar memudahkan bagi reviewer untuk mencermati makna gambar.

|  |
| --- |
| lena%258bit%25plain  Gambar 1 Citra lena.bmp |

## 2. 2 Formulir Copyright

Formulir copyright harus disertakan pada pengiriman naskah akhir. Anda bisa meminta versi .pdf, atau .doc via email ke [indoceiss@gmail.com](mailto:indoceiss@gmail.com)

## 2. 2.1 Rumus Matematika

Jika anda menggunakan *Word,* gunakan persamaan Microsoft Equation Editor atau *MathType*, ditulis ditengah, dan diberi nomor persamaan mulai dari (1), (2) dst.

 (1)

## 2. 2.3 Pengacuan Pustaka

Pengacuan pustaka dilakukan dengan menuliskan [nomor urut pada daftar pustaka] mis. [1], [1,2], [1,2,3]. Sitasi kepustakaan harus ada dalam Daftar Pustaka dan Daftar Pustaka harus ada sitasinya dalam naskah. Pustaka yang disitasi pertama kali pada naskah [1], harus ada pada daftar pustaka no satu, yg disitasi ke dua, muncul pada daftar pustaka no 2, begitu seterusnya. Daftar pustaka urut kemunculan sitasi, bukan urut nama belakang. Daftar pustaka hanya memuat pustaka yang benar benar disitasi pada naskah.

3. HASIL DAN PEMBAHASAN

Pembahasan terhadap hasil penelitian dan pengujian yang diperoleh disajikan dalam bentuk uraian teoritik, baik secara kualitatif maupun kuantitatif. Hasil percobaan sebaiknya ditampilkan dalam berupa grafik atau pun tabel. Untuk grafik dapat mengikuti format untuk diagram dan gambar.



Gambar 2 Grafik perbandingan 

# Tabel 1 Perbandingan Algoritma A dan Algoritma B

|  |  |  |  |
| --- | --- | --- | --- |
| Algoritma | Waktu Proses | Ketelitian | Memori |
| A | 120 ms | 98 % | 200 KB |
| B | 105 ms | 95 % | 415 KB |

4. KESIMPULAN

Kesimpulan harus mengindikasi secara jelas hasil-hasil yang diperoleh, kelebihan dan kekurangannya, serta kemungkinan pengembangan selanjutnya.

Kesimpulan dapat berupa paragraf, namun sebaiknya berbentuk point-point dengan menggunakan numbering atau bullet.

5. SARAN

Saran-saran untuk untuk penelitian lebih lanjut untuk menutup kekurangan penelitian. Tidak memuat saran-saran diluar untuk penelitian lanjut.

# UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada xxx yang telah memberi dukungan financial terhadap penelitian ini.

DAFTAR PUSTAKA

● **Buku** dengan urutan penulisan: Penulis, tahun, *judul buku* (harus ditulis miring) volume (jika ada), edisi (jika ada), nama penerbit dan kota penerbit .

[1] Castleman, K. R., 2004, *Digital Image Processing*, Vol. 1, Ed.2, Prentice Hall, New Jersey.

● **Buku Terjemahan** dengan urutan penulisan: Penulis asli (nama depan, tengah. (disingkat), belakang. (disingkat)), tahun buku terjemahan, *judul bukuterjemahan* (harus ditulis miring), volume (jika ada), edisi (jika ada), (diterjemahkan oleh : nama penerjemah), nama penerbit terjemahan dan kota penerbit terjemahan.

[2] Gonzales, R., P. 2004, *Digital Image Processing* (*Pemrosesan Citra Digital*), Vol. 1, Ed.2, diterjemahkan oleh Handayani, S., Andri Offset, Yogyakarta.

● **Artikel dalam Buku** dengan urutan penulisan: Penulis artikel, tahun, *judul artikel* (harus ditulis miring)*,* nama editor, *judul buku* (harus ditulis miring)*,* volume (jika ada), edisi (jika ada), nama penerbit dan kota penerbit.

[3] Wyatt, J. C, dan Spiegelhalter, D., 1991, *Field Trials of Medical Decision-Aids: Potential Problems and Solutions*, Clayton, P. (ed.): *Proc. 15th Symposium on Computer Applications in Medical Care*, Vol 1, Ed. 2, McGraw Hill Inc, New York.

●***Pustaka dalam bentuk artikel dalam majalah ilmiah:***

Urutan penulisan: Penulis, tahun, judul artikel, *nama majalah* (harus ditulis miring sebagai singkatan resminya), nomor, volume dan halaman.

[4] Yusoff, M, Rahman, S.,A., Mutalib, S., and Mohammed, A. , 2006, Diagnosing Application Development for Skin Disease Using Backpropagation Neural Network Technique, *Journal of Information Technology*, vol 18, hal 152-159.

.

● ***Pustaka dalam bentuk artikel dalam seminar ilmiah:***

**Artikel dalam prosiding seminar** dengan urutan penulisan**:** Penulis, tahun, judul artikel, *Judul prosiding Seminar* (harus ditulis miring), kota seminar, tanggal seminar.

[5] Wyatt, J. C, Spiegelhalter, D, 2008, Field Trials of Medical Decision-Aids: Potential Problems and Solutions, *Proceeding of 15th Symposium on Computer Applications in Medical Care*, Washington, May 3.

● ***Pustaka dalam bentuk Skripsi/Tesis/Disertasi*** dengan urutan penulisan: Penulis, tahun, judul skripsi, *Skipsi/Tesis/Disertasi* (harus ditulis miring), nama fakultas/ program pasca sarjana, universitas, dan kota.

[6] Prasetya, E., 2006, Case Based Reasoning untuk mengidentifikasi kerusakan bangunan, *Tesis,* Program Pasca Sarjana Ilmu Komputer, Univ*.* Gadjah Mada, Yogyakarta.

●***Pustaka dalam bentuk Laporan Penelitian:***

Urutan penulisan: Peneliti, tahun, judul laporan penelitian, *nama laporan penelitian* (harus

ditulis miring), nama proyek penelitian, nama institusi, dan kota.

[7] Ivan, A.H., 2005, Desain target optimal, *Laporan Penelitian Hibah Bersaing,*Proyek Multitahun, Dikti, Jakarta.

***Pustaka dalam bentuk artikel dalam internet*** *(tidak diperkenankan*

*melakukan sitasi artikel dari internet yang tidak ada nama penulisnya)****:***

● **Artikel majalah ilmiah versi cetakan** dengan urutan penulisan: Penulis, tahun, judul artikel, *nama majalah* (harus ditulis miring sebagai singkatan resminya), nomor, volume dan halaman.

[8] Wallace, V. P. , Bamber, J. C. dan Crawford, D. C. 2000. Classification of reflectance spectra from pigmented skin lesions, a comparison of multivariate discriminate analysis and artificial neural network. *Journal Physical Medical Biology* , No.45, Vol.3, 2859-2871.

● **Artikel majalah ilmiah versi *online*** dengan urutan penulisan:Penulis, tahun, judul artikel, *nama majalah* ((harus ditulis miring sebagai singkatan resminya), nomor, volume, halaman dan

alamat website.

[9] Xavier Pi-Sunyer, F., Becker, C., Bouchard, R.A., Carleton, G. A., Colditz, W., Dietz, J., Foreyt, R. Garrison, S., Grundy, B. C., 1998, Clinical Guidlines on the identification, evaluation, and treatment of overweight and obesity in adults, *Journal of National Institutes of Health*, No.3, Vol.4, 123-130, :http://journals.lww.com/acsm-msse/Abstract/1998/11001/paper\_treatment\_of\_obesity.pdf.

● **Artikel umum** dengan urutan penulisan: Penulis, tahun, judul artikel, *alamat website* (harus ditulis miring), diakses tanggal …

[10] Borglet, C, 2003,Finding Asscociation Rules with Apriori Algorithm,[http://www.fuzzy.cs.uniagdeburg](http://www.fuzzy.cs.uni-magdeburg)de/~borglet/apriori.pdf, diakses tgl 23 Februari 2007.

**Daftar Pustaka hanya memuat semua pustaka yang diacu pada naskah tulisan, bukan sekedar pustaka yang didaftar. Pustaka ditulis urut kemunculan pengacuan di naskah, bukan urut abjad penulis.**

[1] Castleman, Kenneth R., 2004, *Digital Image Processing*, Vol. 1, Ed.2, Prentice Hall, New Jersey.

[2] Gonzales, R., P. 2004, *Digital Image Processing* (*Pemrosesan Citra Digital*), Vol. 1, Ed.2, diterjemahkan oleh Handayani, S., Andri Offset, Yogyakarta.

[3] Wyatt, J. C, dan Spiegelhalter, D., 1991, *Field Trials of Medical Decision-Aids: Potential Problems and Solutions*, Clayton, P. (ed.): *Proc. 15th Symposium on Computer Applications in Medical Care*, Vol 1, Ed. 2, McGraw Hill Inc, New York.

[4] Yusoff, M, Rahman, S.,A., Mutalib, S., and Mohammed, A. , 2006, Diagnosing Application Development for Skin Disease Using Backpropagation Neural Network Technique, *Journal of Information Technology*, vol 18, hal 152-159.

[5] Wyatt, J. C, Spiegelhalter, D, 2008, Field Trials of Medical Decision-Aids: Potential Problems and Solutions, *Proceeding of 15th Symposium on Computer Applications in Medical Care*, Washington, May 3.

[6] Prasetya, E., 2006, Case Based Reasoning untuk mengidentifikasi kerusakan bangunan, *Tesis,* Program Pasca Sarjana Ilmu Komputer, Univ*.* Gadjah Mada, Yogyakarta.

[7] Ivan, A.H., 2005, Desain target optimal, *Laporan Penelitian Hibah Bersaing,*Proyek Multitahun, Dikti, Jakarta.

[8] Wallace, V. P. , Bamber, J. C. dan Crawford, D. C. 2000. Classification of reflectance spectra from pigmented skin lesions, a comparison of multivariate discriminate analysis and artificial neural network. *Journal Physical Medical Biology* , No.45, Vol.3, 2859-2871.

[9] Xavier Pi-Sunyer, F., Becker, C., Bouchard, R.A., Carleton, G. A., Colditz, W., Dietz, J., Foreyt, R. Garrison, S., Grundy, B. C., 1998, Clinical Guidlines on the identification, evaluation, and treatment of overweight and obesity in adults, *Journal of National Institutes of Health*, No.3, Vol.4, 123-130, :http://journals.lww.com/acsm-msse/Abstract/1998/11001/paper\_treatment\_of\_obesity.pdf.

[10] Borglet, C, 2003,Finding Asscociation Rules with Apriori Algorithm,[http://www.fuzzy.cs.uniagdeburg](http://www.fuzzy.cs.uni-magdeburg)de/~borglet/apriori.pdf, diakses tgl 23 Februari 2007.