

Analisis Sentimen pada Opini Mahasiswa Menggunakan *Natural Language Processing*

Aldi Nurzahputra¹, Much Aziz Muslim²

^{1,2}Jurusan Ilmu Komputer, FMIPA, Universitas Negeri Semarang
Email: ¹aldinurzah96@gmail.com, ²a212muslim@yahoo.co.id

Abstrak

Dalam peningkatan kualitas pendidikan di Universitas, opini mahasiswa tidak dapat lepas dalam prosesnya. Opini mahasiswa dapat memberikan dukungan terkait proses kegiatan pembelajaran di Universitas. Opini seluruh mahasiswa yang ditampung merupakan data teks yang sangat besar dan tidak terstruktur. Dalam pengolahannya diperlukan teknik analisis sentimen dalam menentukan opini mahasiswa. Teknik analisis sentimen saat ini memiliki peran yang sangat penting. Penggunaan analisis sentimen dapat menentukan sentimen opini dari teks. Penentuan tersebut dapat dilakukan dengan menggunakan teks *mining*. Salah satu penerapan teks *mining* ialah *Opinion mining*. *Opinion mining* dapat melakukan analisis opini dari suatu dokumen teks. Pada penelitian ini membahas *Opinion mining* dari data mahasiswa Ilmu Komputer sebanyak 150 opini mahasiswa menggunakan *Algoritma Naïve Bayes Classification* untuk menganalisis opini mahasiswa. NBC menyelesaikan *Opinion mining* dengan *Part-of-Speech (POS) Tagging* yang digunakan untuk memberikan kelas kata (*tag*) secara gramatikal ke setiap kata dalam suatu kalimat teks. Hasil proses tersebut kemudian diterapkan *rule* untuk mengetahui apakah suatu dokumen termasuk opini atau bukan. Kemudian opini diklasifikasikan ke dalam opini positif dan negatif (*subproses Opinion orientation*). Dari pengujian didapatkan klasifikasi 116 opini positif dan sebanyak 31 opini negatif

Kata Kunci: *Text mining*, opini, POS tagging

Abstract

In improving the quality of education at the University, student Opinion can not be separated in the process. Opinion students can provide support and feedback regarding the process of learning activities at the University. Opinion all students are accommodated a very large Text data and not terstruktur. In pengolahannya sentiment analysis techniques are required in determining student Opinion. Sentiment analysis techniques currently has a very important role. The use of sentiment analysis can determine the sentiment of Opinion Text. The determination can be done using Text mining. One application of Text mining is the Opinion mining. Opinion Opinion mining can perform analysis of a Text document. In this study discusses the Opinion mining of Computer Science student data 150 using Naive Bayes Classification Algorithm to analyze student Opinion. NBC finish Opinion mining with Part-of-Speech (POS) Tagging is used to provide word class (tag) grammatically to each word in a sentence of Text. The result of this process is then applied the rule to determine whether a document including Opinions or not. Then, Opinion classified into positive and negative Opinion (Opinion subprocesses orientation). Classification of the test obtained 116 positive Opinion and as much as 31 negative Opinions

Keyword: *Text mining*, opinion, POS tagging

1. PENDAHULUAN

Peningkatan mutu pendidikan di perguruan tinggi dapat dilakukan dengan melakukan proses evaluasi. Pada dasarnya tujuan evaluasi adalah memperoleh data yang akurat dan objektif tentang pelaksanaan suatu kegiatan. Informasi tersebut dapat mengenai dampak, atau hasil yang dicapai, proses, efisiensi atau pemanfaatan pendayagunaan sumber daya. Dalam prosesnya, penampungan opini civitas akademika menjadi hal yang utama begitu pula opini mahasiswa. Penampungan opini mahasiswa terkait topik tertentu akan menimbulkan beragam opini yang berbeda-beda dengan jumlah yang banyak. Dalam pengolahannya diperlukan teknik analisis sentimen dalam menentukan opini mahasiswa. Analisis sentimen atau *Opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini [1].

Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen [2].

Opinion mining dilakukan untuk melihat pendapat atau kecenderungan opini terhadap masalah atau topik berita oleh seseorang, apakah cenderung beropini negatif, positif atau netral, sehingga diharapkan opini-opini yang terkumpul dapat menjadi informasi yang berguna. Informasi yang terdapat pada berita online merupakan informasi data teks digital yang tidak terstruktur [3]. *Text mining* sangat diperlukan untuk menangani teks yang tidak terstruktur tersebut. *Text mining* mengacu pada proses mengambil informasi berkualitas tinggi dari teks. Kegiatan penting dalam *Text mining* adalah klasifikasi atau kategorisasi. Metode klasifikasi yang digunakan dalam penelitian ini adalah metode *Naïve Bayes* yang sering disebut *Naive Bayes Classifier* (NBC). Eksperimen untuk data review film dan menghasilkan tingkat akurasi 80% menggunakan metode NBC [4]. Penelitian tersebut menunjukkan NBC merupakan metode yang konvensional dan lebih sederhana dari metode lain dalam proses teks *mining* [5]. NBC memberikan hasil yang tepat dalam mengklasifikasikan opini dalam bentuk paragraf yang terdiri dari beberapa kalimat menjadi kesimpulan opini positif, negatif dari opini-opini mahasiswa. Tujuan dari penelitian ini adalah menganalisis sentimen opini mahasiswa dengan dan mengklasifikasi opini mahasiswa dengan metode *Naive Bayes*.

2. METODE

Adapun metode penelitian disini dapat mencakup teori-teori yang digunakan pada tinjauan pustaka yang didapatkan pada literatur dan harus disertai dengan referensi. Informasikan secara ringkas mengenai metode penelitian yang digunakan, jelaskan bagaimana tahapannya. Menurut Turney, *Opinion mining* atau *sentiment analysis* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek yang dikemukakan oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif [6]. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat [7]. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen. *Opinion mining* bisa dianggap sebagai kombinasi antara *Text mining* dan natural language processing. Salah satu metode *Text mining* yang bisa digunakan untuk menyelesaikan masalah *Opinion mining* adalah *Naive Bayes Classifier* (NBC). NBC bisa digunakan untuk mengklasifikasikan opini ke dalam opini positif atau negatif. NBC bisa berfungsi dengan baik sebagai metode pengklasifikasi teks.

Text mining memiliki definisi menambang data yang berupa teks di mana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen. *Text mining* merupakan proses penemuan akan informasi atau trend baru yang sebelumnya tidak terungkap dengan memproses dan menganalisis data dalam jumlah besar. *Text mining* dapat memberikan solusi dari permasalahan seperti pemrosesan, pengorganisasian atau pengelompokkan, dan menganalisis *unstructured Text* dalam jumlah besar. Dalam memberikan solusi, *Text mining* mengadopsi dan mengembangkan banyak teknik dari bidang lain, seperti data *mining*, *information retrieval*, statistika dan matematika, *machine learning*, *linguistic*, *natural language processing*, dan *visualization* [8]. Kegiatan penelitian untuk *Text mining* antara lain ekstraksi dan penyimpanan teks, *preprocessing* akan konten teks, pengumpulan data statistik dan *indexing*, dan analisis konten.

Algoritma *Naive Bayes Classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat [4]. Dalam penelitian ini yang menjadi data uji adalah dokumen opini. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya [9]. Dalam algoritma *Naive Bayes Classifier* setiap dokumen direpresentasikan dengan pasangan atribut " $x_1, x_2, x_3, \dots, x_n$ " dimana x_1 adalah kata pertama, x_2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori opini. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (V_{map}), dimana persamaanya adalah sebagai berikut:

$$V_{MAP} = \underset{v_j \in v}{\arg \max} \left(\frac{P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)} \right) \quad (1)$$

Untuk $P(x_1, x_2, x_3, \dots, x_n)$ nilainya konstan untuk semua kategori (V_j) sehingga persamaan dapat ditulis sebagai Persamaan (2).

$$V_{MAP} = \underset{v_j \in v}{arg\ max} (P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)) \quad (2)$$

Persamaan diatas dapat disederhanakan menjadi Persamaan (3).

$$V_{MAP} = \underset{v_j \in v}{arg\ max} \prod_{i=1}^n (P(x_i | V_j) P(V_j)) \quad (3)$$

Keterangan:

- V_j = Kategori opini
- $P(x_i | V_j)$ = Probabilitas x_i pada kategori V_j
- $P(V_j)$ = Probabilitas dari V_j

Untuk $P(V_j)$ dan $P(x_i | V_j)$ dihitung saat pelatihan sesuai Persamaan (4).

$$P(V_j) = \frac{|docs\ j|}{|contoh|} \quad (4)$$

$$P(x_i | V_j) = \frac{n_k + 1}{n + |kosakata|} \quad (5)$$

Tahap pengolahan awal adalah proses *Text preprocessing* terhadap baris-baris kalimat opini. Tahap ini terdiri dari beberapa fase yaitu tokenisasi, *filtering*, dan *stemming*. Tokenisasi adalah memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata. Tokenisasi menghilangkan delimiter seperti tanda titik (.), koma (,), spasi, dan karakter angka yang ada pada kata tersebut. Tokenisasi juga melakukan fungsi *toLowerCase*, yaitu mengubah semua karakter huruf menjadi huruf kecil. *Filtering* merupakan fase menghilangkan kata-kata yang tidak mengandung makna atau *stopword*. *Stopword* adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen. Sedangkan *stemming* adalah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk kata dasarnya. Tujuan dari proses *stemming* adalah menghilangkan imbuhan-imbuhan baik itu berupa prefiks, sufiks, maupun konfiks yang ada pada setiap kata. Setelah melewati tahapan *preprocessing*, proses selanjutnya adalah membuat model *probabilistic*. Tahapan ini digunakan untuk menghasilkan model yang nilainya akan digunakan pada tahapan selanjutnya yaitu proses klasifikasi.

3. HASIL DAN PEMBAHASAN

Pada subproses ini digunakan metode POS *Tagging* dan *rule* seperti pada subproses *document subjectivity*. Perbedaanya terletak pada susunan *rule* yang digunakan, yaitu aturan kebahasaan untuk menentukan objek yang menjadi target opini. Proses intinya adalah POS *Tagging* menggunakan HMM dan klasifikasi orientasi opini menggunakan NBC. Baik HMM *based POS Tagging* maupun NBC, keduanya diawali oleh proses training menggunakan *training dataset*. POS *tagging* dimuat pada Tabel 1.

Tabel 1. POS *tagging*

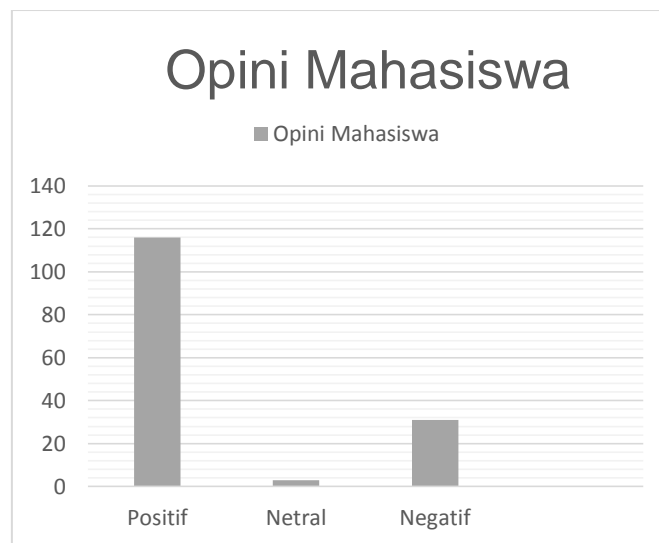
POS Tag	Arti	Contoh
;	Titik Koma	;
:	Titik Dua	:
.	Tanda Titik	.
,	Tanda Koma	,
JJ	Kata Sifat	Baik, Bagus
RB	Kata Keterangan	Sementara, Nanti
NN	Kata Benda	Kursi, Kipas
VBI	Kata Kerja Intransitif	Pergi
VBT	Kata Kerja Transitif	Membeli
IN	Preposisi	Di, Dari, Ke
CC	Kata Sambung Setara	Dan, Atau, Tetapi
SC	Kata Sambung Tidak Setara	Jika, Antara
UH	<i>Injection</i>	Wah, Aduh, Oi
CDO	Bilangan Berurut	Pertama, Kedua, Ketiga
CDP	Bilangan Pokok	Satu, Dua, Tiga
WP	Kata Tanya	Apa, Siapa, Dimana

Untuk menentukan kalimat yang termasuk opini atau bukan, diperlukan *rule* untuk mengolah data hasil proses POS *Tagging*. *Rule* opini yang digunakan dalam penelitian ini ditunjukkan pada Tabel 2.

Tabel 2. Rule POS tagging

No	Rule	Contoh
1	RB JJ	sangat bagus, dengan bagus, benar-benar bagus, seperti jelek, begitu bagus, demikian bagus, agak bagus, amat bagus, sungguh bagus, terlampau bagus, tentu jelek,
2	RB VB	semoga berjalan, semoga membawa hikmah, seandainya datang, jika memilih
3	NN JJ	ruanganya bagus, labnya rapi, perkataannya halus, sistemnya jelek
4	PRP VBI	saya menyukai, kita suka
5	CK JJ	bagus atau baik, tetapi malas

Pengujian subproses *document subjectivity* dan target *detection* dilakukan dengan menggunakan 150 data teks dalam *tagged dataset (dataset POS tagging)*. Data teks yang digunakan dalam pengujian *document subjectivity* dan target *detection* dikumpulkan dengan menggunakan kuisioner. Hasil klasifikasi opini mahasiswa dapat dilihat pada Gambar 1.



Gambar 1. Hasil klasifikasi opini mahasiswa

4. SIMPULAN

Penelitian ini diimplementasikan kedalam aplikasi *Rapid Miner*. *Rapid Miner* merupakan aplikasi untuk pengolahan data *mining*, *Text mining* dan *predictive analytics*. Berdasarkan hasil data opini dapat dilakukan analisis sentimen yang menklasifikasikan sentimen dalam opini atau bukan. *NBC* dapat berfungsi dengan baik pada subproses *Opinion orientation*. Dari hasil klasifikasi menunjukkan opini positif sebanyak 116 opini dan opini negatif sebanyak 31 opini.

5. REFERENSI

- [1] Liu, B. 2010. *Handbook of Natural Language Processing, chapter Sentiment Analysis and Analysis*, 2nd Edition. Chapman & Hall / CRC Press, New York
- [2] Turney, Peter D. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. presented at the *Association for Computational Linguistics 40 Anniversary Meeting*, New Brunswick, N.J.

- [3] Pang, Bo. Lee, L & Vaithyanathan, S. 2002. Sentiment classification using machine learning techniques. *Proceedings of the 7th Conference on Empirical Methods in Natural Language Processing (EMNLP-02)*. USA, 2002
- [4] Kamaruzaman, S.M., Chowdhury M.R. 2004. Text Categorization using Association Rule and Naive Bayes Classifier. *Asian Journal of Information Technology*. Vol. 3(9):657-665
- [5] Kibriya Ashraf M., Frank Eibe, et al. 2004. Multinomial Naïve Bayes for Text Categorization Revisited. *Australian Joint Conference on Artificial Intelligence* No 17. Australia, 4-6 Desember 2004.
- [6] Femphy Pisceldo, Manurung, R., Adriani, Mirna. 2009. Probabilistic Part-of-Speech Tagging for bahasa Indonesia. *Third International MALINDO Workshop, colocated Event ACLIJCNLP 2009*, Singapore, 1 Agustus 2009.
- [7] Wicaksono, Alfian F dan Purwarianti, Ayu. 2010. HMM Based Part-of-Speech Tagger for Bahasa Indonesia. *Proceeding of the Fourth International MALINDO Workshop (MALINDO2010)*. Jakarta, Agustus 2010.
- [8] Jurafsky, Daniel dan Martin, H. James. 2007. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice-Hall, New Jersey
- [9] Liu, B. 2010. *Handbook of Natural Language Processing, chapter Sentiment Analysis and Analysis, 2nd Edition*. Chapman & Hall/CRC Pres, New York