

Uji Komparasi Sentiment Analysis Pada Opini Alumni Terhadap Perguruan Tinggi

I Komang Dharmendra*¹, Ni Nym Utami Januhari ², I Putu Ramayasa³, I Made Agus Wirahadi Putra⁴

^{1,2,3,4} Program Studi Sistem Informasi, ITB STIKOM Bali, Jalan Raya Puputan No 86 Renon Denpasar

e-mail: *¹dharmendra@stikom-bali.ac.id, ²amik@stikom-bali.ac.id,

³ramayasa@stikom-bali.ac.id, ⁴wirahadi@stikom-bali.ac.id

Abstrak

Opini merupakan bagian penting dalam pengambilan keputusan, sehingga diperlukan kemampuan untuk mendapatkan informasi dari opini. Sentiment Analysis adalah cabang ilmu dari Text mining yang bisa digunakan untuk analisa opini yang berupa teks untuk mengelompokkan opini menjadi 3 jenis opini, yaitu opini positif, opini netral dan opini negative. Support Vector Machine (SVM) merupakan salah satu metode yang banyak diterapkan untuk text mining karena mampu menunjukkan performa yang baik (Styawati and Mustofa, 2019). SVM bekerja dengan sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur berdimensi tinggi. Maximum Entropy adalah algoritma klasifikasi probabilistic yang termasuk dalam kelas model eksponensial, yang didasarkan pada prinsip Entropi Maksimum. Maximum Entropy dapat digunakan untuk memecahkan masalah klasifikasi teks seperti seperti deteksi Bahasa, klasifikasi topik, dan analisis sentimen. Dilakukan pengujian analisa sentimen menggunakan metode Support Vector Machine (SVM) dan Maximum Entropy untuk menguji tingkat akurasi setiap metode dalam melakukan analisa sentimen opini alumni perguruan tinggi. dari hasil pengujian menunjukkan Maximum Entropy memiliki tingkat akurasi yang lebih baik dengan hasil 95,45 %.

Kata kunci : Sentimen analysis, SVM, Maximum Entropy.

Abstract

Opinion is an important part of decision making, so it takes the ability to get information from opinions. Sentiment Analysis is a branch of science from Text mining that can be used for opinion analysis in the form of text to classify opinions into 3 types of opinions, namely positive opinions, neutral opinions and negative opinions. Support Vector Machine (SVM) is one method that is widely applied for text mining because it is able to show good performance (Styawati and Mustofa, 2019). SVM works with a learning system that uses a hypothetical space in the form of linear functions in a high-dimensional feature space. Maximum Entropy is a probabilistic classification algorithm that belongs to the class of exponential models, which is based on the principle of Maximum Entropy. Maximum Entropy can be used to solve text classification problems such as Language detection, topic classification, and sentiment analysis. Sentiment analysis was tested using the Support Vector Machine (SVM) and Maximum Entropy methods to test the accuracy of each method in analyzing the sentiments of college alumni opinions. from the test results show Maximum Entropy has a better level of accuracy with the results of 95.45%.

Keywords : Sentiment analysis, SVM, Maximum Entropy

1. PENDAHULUAN

Dalam proses pengambilan keputusan, diperlukan dasar dan landasan yang kuat agar keputusan yang diambil dapat memberikan dampak yang maksimal dan tidak merugikan perusahaan, di Perguruan Tinggi yang bisa menjadi dasar dalam pengambilan keputusan adalah opini dari para alumni. Opini dari alumni dikumpulkan saat mahasiswa telah menyelesaikan masa studi dan dinyatakan lulus, dan opini yang disampaikan dari mahasiswa tersebut adalah hal yang mereka alami selama menempuh Pendidikan. Untuk mendapatkan informasi dari opini yang disampaikan oleh alumni dilakukan analisa sentimen, untuk mengeluarkan informasi apakah opini yang disampaikan oleh alumni tersebut bernilai positif, negative, atau netral[1].

Analisa sentiment atau *Sentiment Analysis* adalah studi komputasional dari opini-opini orang, *appraisal* dan emosi melalui entitas, event dan atribut yang dimiliki[2], [3]. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau

fitur/tingkat aspek - apakah pendapat yang dikemukakan dalam dokumen, kalimat atau fitur entitas / aspek bersifat positif, negatif atau netral.

Support Vector Machine (SVM) merupakan salah satu metode yang banyak diterapkan untuk berbagai jenis penelitian, termasuk pada bidang data dan text mining karena mampu menunjukkan performa yang baik[4]. SVM bekerja dengan sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur berdimensi tinggi[4]. Pada analisa sentimen berbahasa Indonesia pada data review film menunjukkan *Support Vector Machine SVM* menghasilkan tingkat akurasi yang lebih baik dibandingkan *Naive Bayes (NBC)* [5].

Maximum Entropy adalah algoritma klasifikasi probabilistic yang termasuk dalam kelas model eksponensial[6], yang didasarkan pada prinsip Entropi Maksimum. *Maximum Entropy* dapat digunakan untuk memecahkan masalah klasifikasi teks seperti seperti deteksi Bahasa, klasifikasi topik, dan analisis sentimen[7]. *Maximum Entropy* dapat memanfaatkan data kontinu dan kategoris, dan dapat menggabungkan interaksi antara variabel yang berbeda. Selain itu efisien, karena sudah dikembangkan dan dijamin dapat bertemu dengan distribusi probabilitas optimal. Namun *Maximum Entropy* memiliki kekurangan yaitu performa rendah dengan fitur independen dan memakan memori karena komputasi yang rumit dan banyak[8].

Tujuan penelitian ini adalah melihat melakukan analisa sentimen dari metode *Support Vector Machine (SVM)* dan *Maximum Entropy* dalam melakukan analisa sentiment berbahasa Indonesia dengan tiga kelas klasifikasi, yaitu positif, negatif, dan netral. Dan membandingkan tingkat akurasi dari masing-masing metode.

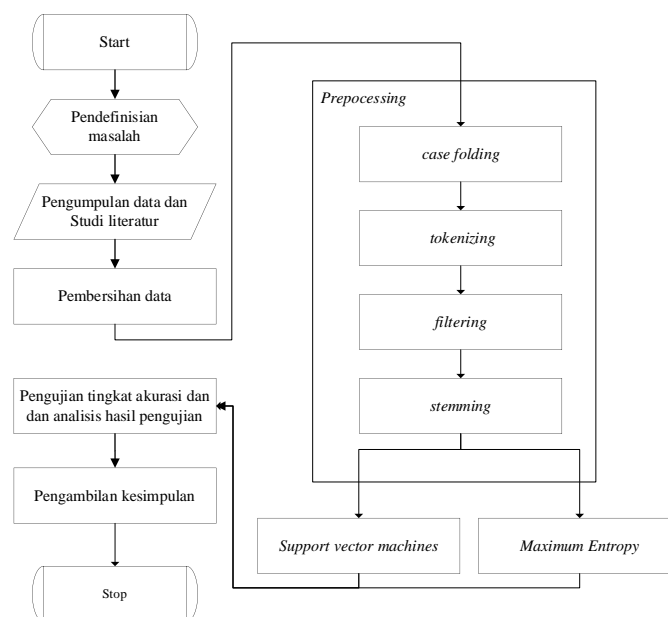
2. METODE PENELITIAN

Pada penelitian ini menggunakan dua metode dalam melakukan analisa sentiment, yaitu metode *Support Vector Machine (SVM)* dan *Maximum Entropy*. Diagram Alur Penelitian dapat dilihat pada gambar 1. Data yang digunakan berjumlah 366 opini alumni yang telah dirapikan dengan mengubah kata yang disingkat penulisannya tanpa merubah makna dan arti dari kalimat tersebut. Data kemudian dibagi menjadi data latih dan data uji, dengan 300 data latih, dan 66 data uji.

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing

Struktur data yang baik dapat memudahkan proses komputerisasi secara otomatis. Dalam penggalan teks, teks dokumen yang digunakan harus dipersiapkan terlebih dahulu sebelum dapat digunakan untuk proses utama. Proses mempersiapkan teks dokumen atau *dataset* mentah disebut dengan proses praproses teks (*text preprocessing*) [9], [10]. Praproses teks ini bertujuan untuk mengubah data teks yang tidak terstruktur menjadi data yang terstruktur[11]. Praproses terdiri dari beberapa tahapan diantaranya *case folding*, *tokenizing*, *filtering* dan *stemming*.



Gambar 1. Alur Penelitian

3.1.1 Case folding

Case folding merupakan tahapan yang mengubah semua huruf dalam dokumen menjadi huruf kecil (*toLowerCase*) dan hanya huruf „a” hingga „z” yang diterima. Karakter selain huruf dihilangkan dan dianggap sebagai delimiter, termasuk angka dan tanda baca[5].

3.1.2 Tokenizing

Tahap *tokenizing* atau disebut juga *parsing* adalah pemrosesan sebuah dokumen menjadi unit kecil yang disebut dengan token, bisa berupa suatu kata, angka, atau tanda baca. Dengan kata lain, *tokenizing* merupakan proses pemotongan string masukan berdasarkan tiap kata yang menyusunnya[7].

Tabel 1. Preprosesing teks opini alumni

Tahapan	Penggunaan
Kalimat Asli	Ruang kelas dalam kondisi baik, dan selalu dibersihkan setelah selesai untuk dipakai oleh kelas selanjutnya
<i>Case Folding</i>	ruang kelas dalam kondisi baik dan selalu dibersihkan setelah selesai untuk dipakai oleh kelas selanjutnya
<i>Tokenizing</i>	ruang kelas dalam kondisi baik dan selalu dibersihkan setelah selesai untuk dipakai oleh kelas selanjutnya
<i>Filtering</i>	ruang kelas kondisi baik dibersihkan setelah selesai dipakai kelas selanjutnya
<i>Stemming</i>	ruang kelas kondisi baik bersih selesai pakai kelas lanjut

3.1.3 Filtering

Tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil *tokenizing*. Proses *filtering* dapat menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata tidak deskriptif yang bukan merupakan kata unik pada suatu dokumen, yang dapat dibuang dalam pendekatan *bag-of-words*. Kata-kata yang termasuk ke dalam *stopword* dihilangkan karena terlalu sering muncul dalam dokumen dan bukan merupakan pembeda yang baik dalam proses penggalan teks[12]. *Stopword* dapat berupa kata depan atau kata sambung, seperti “yang”, “dan”, “di”, “dari”, dan sebagainya [8]

3.1.3 Stemming

Stemming adalah suatu proses pengembalian suatu kata berimbuhan ke dalam bentuk dasarnya. Proses ini akan menghilangkan awalan, akhiran, sisipan dan *confixes* (kombinasi dari awalan dan akhiran). Proses ini memiliki banyak variasi sesuai domain bahasa yang digunakan, karena imbuhan tiap bahasa berbeda-beda. Proses *stemming* pada teks berbahasa Indonesia berbeda dengan *stemming* pada teks berbahasa Inggris. Pada teks berbahasa Inggris, proses yang diperlukan hanya proses menghilangkan sufiks. Sedangkan pada teks berbahasa Indonesia, selain sufiks, prefiks, dan konfiks juga dihilangkan [13].

Stemming yang digunakan adalah stemming Sastrawi, yang merupakan stemmer pengembangan dari Algoritma Nazief dan Adriani yang awalnya dibangun pada Bahasa pemrograman PHP, kemudian dikembangkan pada Bahasa pemrograman python, java, C, Go, Ruby. Sastrawi sangat bergantung pada kamus kata dasar yang diambil dari kateglo.com dengan perubahan[14].

3.2 analisa sentimen

Proses analisa sentimen menggunakan dua metode, metode *Support Vector Machine (SVM)* dan *Maximum Entropy*. Dimana masing-masing metode akan digunakan untuk melakukan analisa sentimen secara terpisah dan dibandingkan tingkat akurasi dari setiap metode.

3.2.1 Support Vector Machine (SVM)

Setelah proses preprosesing data selesai dilakukan, dilanjutkan dengan proses ekstraksi fitur pada data latih untuk pengambilan ciri opini yang dapat menggambarkan karakteristik dari jenis opini yang ada.

Setelah ekstraksi fitur selesai dilakukan maka dilanjutkan dengan proses klasifikasi, dimana data opini akan diproses perbaris dari 66 opini data uji yang telah disiapkan. dengan mengelompokkan hasil prediksi menjadi opini positif jika nilai prediksi adalah 0, opini negative ketika nilai prediksi adalah 1, dan netral jika nilai prediksi diluar dua nilai tersebut.

```

C:\WINDOWS\system32\cmd.exe
D:\Library\Documents\GitHub\sentimentanalysis>python "Support Vector Machine -RAW.py"
Output : positive
Probs : [[0.33333333 0.33333333 0.33333333]]
Done in : 0.057719
Output : negative
Probs : [[0.33968577 0.32678831 0.33352592]]
Done in : 0.0379413
Output : positive
Probs : [[0.3380464 0.32876743 0.33318617]]
Done in : 0.0391542
Output : neutral
Probs : [[0.32337541 0.34283242 0.33379217]]
Done in : 0.0377781
  
```

Gambar 2. Proses analisa sentimen menggunakan *Support Vector Machine (SVM)*

3.2.2 Maximum Entropy

Dalam implementasi metode Maximum Entropy digunakan data yang sama seperti yang digunakan dalam percobaan menggunakan metode SVM. Setelah preprocessing data dilakukan ekstraksi fitur kemudian dilanjutkan klasifikasi opini. Dalam percobaan ini juga digunakan library nltk yang ada pada bahasa pemrograman python. File opini yang digunakan disimpan dalam berkas .csv, dimana untuk data uji diolah setiap baris untuk menemukan opini dari masing-masing kalimat.

```

testSentimentAnalysis = ngDosen memiliki metode belajar yang menyenangkan, sehingga mahasiswa mudah menangkap materi., s
sentiment = negative, time = 0.0022723
testSentimentAnalysis = tugas dan cara mengajar dosen yang sangat baik dan bagus, sentiment = negative, time = 0.0016969
testSentimentAnalysis = Dosen lebih banyak berinteraksi dalam penyampaian materi, dapat memberikan suasana belajar yang
aktif dikelas, sentiment = negative, time = 0.0016784
testSentimentAnalysis = Proses Perkuliahan bagus dengan pengajar yang sesuai dengan bidangnya, sentiment = positive, time
= 0.001609
testSentimentAnalysis = Materi yang diberikan sesuai dengan silabus, sentiment = positive, time = 0.0023439
testSentimentAnalysis = Dosen yang mengajar kompeten dalam memberikan materi, sentiment = positive, time = 0.0015711
testSentimentAnalysis = sudah sangat cukup membantu dan sangat efektif, karena itulah yang mahasiswa butuhkan, yaitu penj
elasan materi secara detail, sentiment = positive, time = 0.0024025
testSentimentAnalysis = Dosen yang bersangkutan dengan materi pembelajaran, sudah memberikan materi dengan maksimal, senti
ment = positive, time = 0.0016324
testSentimentAnalysis = Dosen yang memiliki kreativitas dalam cara mengajar akan meningkatkan keinginan mahasiswa untuk b
  
```

Gambar 3. Proses analisa sentimen menggunakan *Maximum Entropy*

3.3. Pengujian

Dari tahap pengujian dengan menggunakan 300 data latih yang dilakukan maka didapatkan hasil pengujian analisa opini. Tabel 2 merupakan jumlah opini yang berhasil dianalisa dengan menggunakan total 66 opini yang dibagi menjadi 3 katagori, yaitu positif, negatif, dan netral. Untuk menghitung tingkat akurasi sentiment digunakan persamaan berikut:

$$\text{Akurasi Jenis Opini} = \frac{\text{Jumlah data uji yang sesuai}}{\text{jumlah data uji berdasarkan opini}} \times 100\%$$

Pegujian untuk algoritma Maximum Entropy dan SVM didapatkan bahwa proses sentiment analis pada algoritma Maximum Entropy menunjukkan tingkat akurasi tertinggi mencapai 95,45 % dengan 59 opini yang berhasil dianalisa dari 66 opini. Tingkat akurasi terendah dicapai pada algoritma SVM dengan menggunakan stemming sastrawi, dengan tingkat akurasi mencapai 71,21% dengan 47 opini yang berhasil dianalisa dari 66 opini. Gambar 4 menunjukkan tingkat akurasi analisa opini dengan menggunakan 300 data latih.

Tabel 2. Persentase Tingkat Akurasi Analisa Opini Alumni

	positive	negative	neutral	Total
<i>Maximum Entropy</i>	18	20	21	59
<i>SVM</i>	13	17	17	47

Tingkat akurasi secara keseluruhan menunjukkan keunggulan algoritma Maximum Entropy lebih dibandingkan dengan algoritma SVM untuk melakukan Analisa opini dengan 3 jenis opini, yaitu opini positif, negative, dan netral. Mengacu pada table 4.6 persentase akurasi tertinggi pada algoritma Maximum Entropy mencapai 94,45%, sedangkan pada algoritma SVM hanya 71,21% saja.



Tabel 4. Kemunculan 15 kata terbanyak pada opini alumni

No	Kata	Kemunculan
1	dosen	95
2	materi	90
3	mahasiswa	78
4	kuliah	61
5	kelas	59
6	ajar	49
7	ruang	29
8	kampus	25
9	tugas	24
10	sesuai	22
11	terkadang	21
12	komputer	18
13	stikom	18
14	laboratorium	17
15	bersih	17

1. Pengujian untuk algoritma Maximum Entropy menunjukkan tingkat akurasi mencapai 95,45 %.
2. Pengujian untuk algoritma SVM menunjukkan tingkat akurasi mencapai 71,21%.
3. Dari hasil pengujian didapatkan bahwa penggunaan algoritma Maximum Entropy menunjukkan hasil yang lebih baik saat melakukan sentiment analysis Bahasa Indonesia pada opini alumni perguruan tinggi.
4. Pada penelitian berikutnya disarankan untuk menggunakan data yang lebih banyak, mulai dari 1000 data atau 10000 data untuk mendapatkan perbandingan nilai akurasi yang lebih akurat.
5. Pada tahapan pra-pengolahan opini dapat dicoba menggunakan algoritma berbasis kamus, seperti Arifin dan Setiono:.

DAFTAR PUSTAKA

- [1] K. Dharmendra, K. O. Saputra, and I. N. Pramaita, 'Analisa Sentiment Untuk Opini Alumni Perguruan Tinggi', *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 2, pp. xxxx–xxxx, Jul. 2019, doi: 10.24843/mite.2019.v18i02.p11.
- [2] J. Riany, M. Fajar, and M. P. Lukman, 'Penerapan Deep Sentiment Analysis pada Angket Penilaian Terbuka Menggunakan K-Nearest Neighbor', *Open Access J. Inf. Syst.*, vol. 06, no. 01, pp. 147–156, 2016.
- [3] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata, 'SENTIMENT ANALYSIS ON E-SPORTS FOR EDUCATION CURRICULUM USING NAIVE BAYES AND SUPPORT VECTOR MACHINE', *J. Ilmu Komput. Dan Inf.*, vol. 13, no. 2, pp. 109–122, Jul. 2020, doi: 10.21609/jiki.v13i2.885.
- [4] S. Styawati and K. Mustofa, 'A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification', *IJCCS Indones. J. Comput. Cybern. Syst.*, vol. 13, no. 3, pp. 219–230, Jul. 2019, doi: 10.22146/IJCCS.41302.
- [5] E. Indrayuni, 'KOMPARASI ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK ANALISA SENTIMEN REVIEW FILM', *J. Pilar Nusa Mandiri*, vol. 14, no. 2, pp. 175–180, Sep. 2018, doi: 10.33480/PILAR.V14I2.36.
- [6] K. Fragos, K. Fragos, Y. Maistros, and C. Skourlas, 'A Weighted Maximum Entropy Language Model for Text Classification', Accessed: May 22, 2021. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.96.964>
- [7] G. A. Buntoro, 'Sentiments Analysis for Governor of East Java 2018 in Twitter', *Sinkron*, vol. 3, no. 2, p. 49, Dec. 2019, doi: 10.33395/sinkron.v3i2.10025.
- [8] X. Xie, S. Ge, F. Hu, M. Xie, and N. Jiang, 'An improved algorithm for sentiment analysis based on maximum entropy', *Soft Comput.*, vol. 23, no. 2, pp. 599–611, Jan. 2019, doi: 10.1007/S00500-017-2904-0.
- [9] P. R. Thanvi, N. S. Sontakke, S. R. Waghmare, Z. S. Patel, and S. Gavhane, 'Sentiment Analysis for Political Reviews using AAVN Combinations', pp. 72–74, 2017.
- [10] R. Bogdan, 'Sentiment Analysis on Embedded Systems Blended Courses', no. 2, pp. 17–23, 2016.
- [11] F. Rahutomo, Z. Hanif Rachmat Adi, I. Fahrur Rozi, and P. Yoga Saputra, 'Implementasi Text Mining Pada Website/Blog Di Internet Untuk Menilai Kinerja Suatu Organisasi', *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 2, p. 101, Nov. 2018, doi: 10.35314/isi.v3i2.462.
- [12] P. Choudhari, 'Sentiment Analysis and Machine Learning Based Sentiment Classification : A Review', vol. 8, no. 3, 2017.
- [13] I. M. A. Agastya, 'Pengaruh Stemmer Bahasa Indonesia Terhadap Peforma Analisis Sentimen Terjemahan Ulasan Film', *J. Tekno Kompak*, vol. 12, no. 1, pp. 18–23, Feb. 2018, doi: 10.33365/JTK.V12I1.70.
- [14] 'Stemming Bahasa Indonesia · sastrawi/sastrawi Wiki · GitHub'. <https://github.com/sastrawi/sastrawi/wiki/Stemming-Bahasa-Indonesia> (accessed Jun. 26, 2019).