

# ANALISIS SENTIMEN PENGGUNA JEJARING SOSIAL MENGUNAKAN METODE SUPPORT VECTOR MACHINE

M. Fachrurrozi, M.T.<sup>1</sup>, Novi Yusliani, M.T.<sup>2</sup>

<sup>1,2</sup>Jurusan Teknik Informatika  
Fakultas Ilmu Komputer, Universitas Sriwijaya  
<sup>1</sup>obetsobets@gmail.com, <sup>2</sup>novi.yusliani@gmail.com

---

## Abstrak

Jejaring sosial membantu pengguna Internet dalam berkomunikasi. Hal ini dikarenakan pengguna jejaring sosial dapat menyampaikan pesan dengan memanfaatkan fasilitas yang disediakan oleh setiap jejaring sosial. Pesan yang akan disampaikan melalui jejaring sosial dapat di-post melalui antarmuka situs web jejaring sosial, pesan singkat (SMS), atau melalui berbagai aplikasi untuk perangkat seluler. Pesan-pesan para pengguna jejaring sosial dapat dimanfaatkan untuk berbagai hal, seperti melakukan review terhadap suatu produk atau melakukan review terhadap suatu masalah baik politik maupun masalah sosial terkini. Hal ini dapat dilakukan dengan menganalisis sentimen para pengguna yang disampaikan melalui pesan-pesan dalam situs jejaring sosial. Metode Support Vector Machine merupakan salah satu metode yang dapat dimanfaatkan untuk menganalisis sentimen. Analisis sentimen dengan menggunakan metode Support Vector Machine dilakukan dengan cara mengklasifikasikan sentimen ke dalam kelas positif atau kelas negatif. Tingkat akurasi sistem analisis sentimen pengguna jejaring sosial dengan menggunakan metode Support Vector Machine adalah 81 % atau sebanyak 243 kicauan yang terkelompok secara tepat dari jumlah keseluruhan kicauan sebanyak 300 kicauan.

**Kata kunci :** *analisis sentimen, jejaring sosial, support vector machine*

---

## 1. Pendahuluan

Jejaring sosial saat ini telah menguasai kehidupan para pengguna Internet. Layanan yang diberikan oleh setiap situs jejaring sosial berbeda-beda. Hal ini menandakan bahwa setiap situs jejaring sosial memiliki ciri khas dan keunggulan tersendiri. Pada umumnya, layanan yang ada pada jejaring sosial adalah chatting, e-mail, berbagi pesan, berbagi video atau foto, forum diskusi, blog, dan lain-lain. Pemanfaatan situs jejaring sosial telah menjadi trend atau gaya hidup sebagian masyarakat.

Jejaring sosial membantu pengguna Internet dalam berkomunikasi. Hal ini dikarenakan pengguna jejaring sosial dapat menyampaikan pesan dengan memanfaatkan fasilitas yang disediakan oleh setiap jejaring sosial. Pesan yang akan disampaikan melalui jejaring sosial dapat di-post melalui antarmuka situs web jejaring sosial, pesan singkat (SMS), atau melalui berbagai aplikasi untuk perangkat seluler. Dengan demikian, pesan-pesan para pengguna jejaring sosial dapat dimanfaatkan untuk berbagai hal, seperti melakukan review terhadap suatu produk atau melakukan review terhadap suatu masalah baik politik maupun masalah

sosial terkini. Hal ini dapat dilakukan dengan menganalisis sentimen para pengguna yang disampaikan melalui pesan-pesan dalam situs jejaring sosial.

Analisis sentimen merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi (Bo Pang dan Lillian, 2008) [1]. Hal ini menandakan bahwa analisis sentimen dapat digunakan untuk mendapatkan informasi yang ada di dalam pesan para pengguna jejaring sosial terhadap topik yang dibahas.

Analisis sentimen dapat dilakukan dengan mengklasifikasi informasi yang ada ke dalam kelas sentimen positif dan ke dalam kelas sentimen negatif. Informasi akan diklasifikasikan ke dalam kelas positif apabila informasi yang disampaikan bernilai baik atau setuju terhadap sesuatu. Sebaliknya, informasi akan diklasifikasikan ke dalam kelas negatif apabila informasi yang disampaikan bernilai tidak baik atau tidak setuju. Penelitian ini bertujuan untuk menganalisis sentimen pengguna jejaring sosial terhadap suatu topik bahasan yang ada dengan memanfaatkan pesan para pengguna jejaring sosial.

## 2. Dasar Teori

### A. Support Vector Machine

Support Vector Machine (SVM) adalah sistem pembelajaran mesin yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (feature space) berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan learning bias yang berasal dari teori pembelajaran statistik (Christianini, et.al., 2000) [2].

Menurut (yuhariz, 2013)[3] SVM adalah sistem pembelajaran mesin yang bekerja atas prinsip *Structural Risk Minimization* (SRM). SRM bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah kelas pada *input space*. *Hyperplane* terbaik antara dua kelas dapat ditemukan dengan mengukur *margin hyperplane* dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperplane* dengan *pattern* terdekat dari masing-masing kelas. *Pattern* yang paling dekat disebut sebagai *support vector*.

Tahapan pembelajaran dalam SVM adalah :

#### 1. Mencari Lagrange Multipliers ( $\alpha$ )

$$\bar{L}(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (1)$$

dikenakan (untuk setiap  $i = 1, \dots, n$ )

Keterangan:

$y_i$ : kelas data latih (+1/-1).

$y_j$ : kelas data latih (+1/-1).

$x_i$ : vektor bobot kalimat komentar.

$x_j$ : vektor bobot kalimat komentar.

#### 2. Mencari Nilai Bobot ( $w$ )

$$w = \sum_{i=1}^n (\alpha_i y_i x_i) \quad (2)$$

Keterangan :

$w$ : vektor bobot.

$y_i$ : kelas data latih (+1/-1).

$x_i$ : vektor bobot kalimat komentar yang menjadi vektor pendukung.

#### 3. Mencari Nilai Bias ( $b$ )

$$b = \frac{1}{NSV} \sum_{i=1}^{NSV} (w \cdot x_i - y_i) \quad (3)$$

Keterangan :

NSV : jumlah vektor pendukung.

$w$ : vektor bobot

$y_i$ : kelas data latih (+1/-1).

$x_i$ : vektor bobot kalimat komentar yang menjadi vektor pendukung.

Proses pengklasifikasian (pengujian) dalam SVM menggunakan persamaan 4.

$$f(\vec{t}) = \text{sgn} \left( \sum_{i=1, x_i \in SV}^n \alpha_i y_i < \vec{t} \cdot \vec{x}_i > + b \right) \quad (4)$$

Keterangan :

$t$ : vektor bobot data uji

$x_i$ : vektor pendukung

$b$ : nilai bias

$y_i$ : kelas atau label dari vektor pendukung (+1/-1)

### B. Analisis Sentimen

Analisis sentimen merupakan salah satu cabang penelitian di domain *Text Mining*. Secara umum, analisis sentimen berkaitan dengan riset komputasional dari opini, sentimen, dan emosi yang diekspresikan secara tekstual. Analisis sentimen bertujuan untuk mengekstrak atribut dari sebuah komentar (opini, sentimen, dan emosi) yang diekspresikan secara tekstual. Setelah itu, akan diberikan penilaian terhadap komentar tersebut, yaitu positif atau negatif (Bo Pang dan Lillian, 2008) [1]. Secara umum, analisis sentimen dibagi menjadi dua kategori, yaitu:

a. *Coarse-grained sentiment analysis*

b. *Fined-grained sentiment analysis*

*Coarse-grained sentiment analysis* melakukan proses analisis pada level dokumen. Pengklasifikasian berorientasi pada sebuah dokumen secara keseluruhan, yaitu positif, netral, dan negatif. Sedangkan *fined-grained sentiment analysis* melakukan proses analisis sebuah kalimat. Tabel 1 menunjukkan contoh *fined-grained sentiment analysis*.

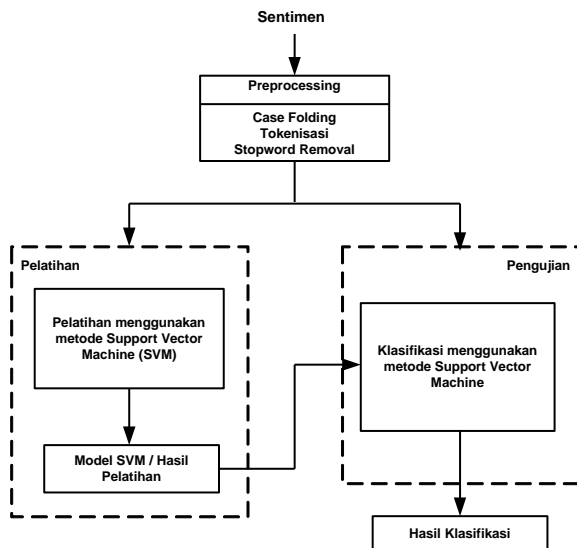
Tabel 1. Contoh *Fined-Grained Sentiment Analysis*

Kalimat	Klasifikasi
Saya tidak suka <i>programming</i>	Negatif
Hotel itu sangat indah sekali	Positif

### 3. Sistem Analisis Sentimen menggunakan metode *Support Vector Machine*

Sistem analisis sentimen adalah sistem yang berfungsi untuk menganalisis sentimen para pengguna jejaring sosial. Analisis sentimen dapat

dilakukan dengan mengklasifikasi informasi yang ada di dalam komentar atau kicauan para pengguna jejaring sosial ke dalam kelas sentimen positif dan ke dalam kelas sentimen negatif. Informasi akan diklasifikasikan ke dalam kelas positif apabila informasi yang disampaikan bernilai baik atau setuju terhadap sesuatu. Sebaliknya, informasi akan diklasifikasikan ke dalam kelas negatif apabila informasi yang disampaikan bernilai tidak baik atau tidak setuju.



Gambar 1. Diagram Proses Sistem Analisis Sentimen

Gambar 1 menggambarkan diagram proses dari sistem analisis sentimen menggunakan metode *Support Vector Machine*. Sistem ini terdiri tiga proses utama yaitu *Pre-processing*, Pelatihan, dan Pengujian.

#### A. Pre-Processing

*Pre-processing* (pemrosesan awal dokumen) merupakan tahapan yang berfungsi untuk mentransformasikan dokumen ke dalam bentuk representasi yang lain. Tujuan dari tahap ini adalah untuk mempercepat pemrosesan terhadap dokumen, mempermudah proses pencarian query ke dalam dokumen, dan mempermudah proses pengurutan dokumen-dokumen yang terambil (*retrieved*) (yusliani, 2010). Proses-proses yang dilakukan dalam *pre-processing* yaitu *casefolding*, *stopword removal*, dan tokenisasi.

*Casefolding* adalah proses penyeragaman bentuk huruf dalam sebuah dokumen, yaitu menjadikan semua huruf ke dalam bentuk huruf kecil. Hanya huruf ‘a’ sampai dengan ‘z’ yang diterima. Karakter selain huruf dianggap delimiter. Tokenisasi adalah proses memecah kalimat menjadi kumpulan kata. Pemecahan dilakukan berdasarkan spasi. *Stopwords removal* adalah kata-kata yang muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh stopwords dalam bahasa Indonesia diantaranya

‘yang’, ‘ke’. Daftar *stopwords* disesuaikan dengan kebutuhan atau kasus. Tabel 2 menunjukkan contoh proses-proses dalam *pre-processing*.

Tabel 2. Contoh Proses-Proses dalam *Pre-Processing*

<i>Pre-Processing</i> Kalimat	
Kalimat	Tuti membeli sepeda di Palembang
<i>Casefolding</i>	tuti membeli sepeda di palembang
<i>Tokenizing</i>	“tuti” “membeli” “sepeda” “di” “palembang”
<i>Stopwords removal</i>	tuti membeli sepeda palembang

#### B. Pelatihan

Pelatihan adalah proses yang dilakukan untuk mendapatkan pola sentimen dari setiap kelas yaitu kelas positif dan kelas negatif. *Support Vector Machine* adalah salah satu metode yang membutuhkan pelatihan sebelum dilakukan proses pengklasifikasian (pengujian).

#### C. Pengujian

#### 4. Pengujian

Pengujian dilakukan dengan menggunakan kicauan pengguna di salah satu jejaring sosial. Pengelompokan dilakukan dengan memilih topik yang ingin dikelompokkan. Jumlah kicauan yang diujikan berjumlah 300 sampel kicauan yang diujikan kepada 7 mahasiswa. Hasil pengelompokan kicauan ini akan dibandingkan dengan pilihan responden untuk mengetahui presisi hasil klasifikasi dengan pikiran manusia untuk menentukan kicauan positif dan negatif.

Responen akan memilih benar atau salah dalam menentukan hasil dari pengujian ini setelah didapat hasil dari 7 responen maka akan dihitung mana yang paling banyak apakah satu kicauan tersebut lebih banyak benarnya atau salahnya apabila lebih banyak benarnya maka kicauan tersebut dianggap benar masuk dalam kelompoknya apabila salah maka kicauan tersebut dianggap salah. Akurasi sistem didapat dengan menggunakan persamaan 5.

$$\frac{\text{Jumlah Klasifikasi Benar}}{\text{Jumlah Seluruh Data Uji}} \times 100\% \quad (5)$$

Hasil pengujian tingkat akurasi sistem memberikan hasil yang optimal, jika dihitung secara keseluruhan, hanya sekitar 81 % atau sebanyak 243

kicauan yang terkelompok secara tepat dari jumlah keseluruhan kicauan sebanyak 300 kicauan. Data pembandingan merupakan data yang diperoleh dari responden dengan memberikan mereka kuisisioner untuk diisi.

Persentase akurasi sebesar 81% didapatkan dari kuisisioner yang dilakukan kepada 7 mahasiswa, dengan jumlah pengelompokan yang benar didapatkan dari jumlah terbanyak yang sama dengan hasil pengelompokan perangkat lunak. Kuisisioner dilakukan dengan cara membandingkan hasil yang dikelompokkan oleh perangkat lunak dengan opini seseorang terhadap kicauan tersebut.

Data latih yang digunakan pada pengujian diatas sebanyak masing-masing 120 kicauan, baik untuk kicauan positif maupun kicauan negatif. Dengan 120 kicauan data latih didapatkanlah hasil yang optimal dalam akurasi sistem. Pemilihan data latih dapat mempengaruhi hasil dari akurasi sistem yang dibangun. Data latih yang dipilih dengan baik akan meningkatkan akurasi sistem. Kicauan yang dianggap tepat sebagai data latih haruslah diuji terlebih dahulu. Pengujian kicauan yang digunakan dalam data latih berupa menanyakan secara langsung kepada ahli bahasa apakah kicauan tersebut tepat dikatakan positif atau negatif.

Pada pengujian ini juga dihitung waktu yang dibutuhkan perangkat lunak untuk memberikan review terhadap suatu topik tersebut. Waktu yang dibutuhkan untuk menangkap satu kicauan kurang lebih 1 detik maka apabila kicauan yang ingin diambil berjumlah 300 maka waktu yang dibutuhkan adalah 300 detik atau sekitar 5 menit dengan kecepatan internet 0.82 Mbps. Kecepatan internet mempengaruhi hasil dari kecepatan proses pada perangkat lunak karena perangkat lunak bekerja secara online. Apabila kecepatan internet lebih cepat maka waktu yang dibutuhkan akan cepat pula.

## 5. Kesimpulan

Sistem analisis sentimen pengguna jejaring sosial dengan menggunakan metode Support Vector Machine menghasilkan tingkat akurasi sebesar 81%. Akurasi sistem salah satunya dipengaruhi oleh jumlah Data Latih yang digunakan dalam proses pelatihan.

Kesalahan klasifikasi sering terjadi pada kicauan yang memiliki makna ganda atau ambigu. Kicauan yang memiliki makna ganda atau ambigu dapat dipandang sebagai komentar positif dan dapat pula dipandang sebagai komentar negatif oleh responden, sehingga sistem juga terkadang bisa salah klasifikasi.

Peningkatan akurasi sistem dapat dilakukan dengan memilih data latih yang benar secara teliti. Selain itu, sebaiknya kicauan atau komentar para pengguna jejaring sosial dicek terlebih dahulu. Hal ini dikarenakan sering kali pengguna jejaring sosial

menulis kicauan yang tidak sesuai dengan tata bahasa Bahasa Indonesia yang baik.

## Daftar Pustaka:

- [1] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.
- [2] Christianini, N., & Shawe-Taylor, J. (2000). *Support vector machines*. Cambridge, UK: Cambridge University Press, 93(443), 935-948.
- [3] Aldyan Y. (2013) Analisis persepsi dan sentimen berita melalui komentar pada artikel berita online dengan pendekatan Support Vector Machine (SVM).

