

IMPLEMENTASI METODE MULTI CLASS SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI EMOSI PADA LIRIK LAGU BAHASA INDONESIA

**TRY OUT SEMINAR SEMESTER GENAP
2016/2017**

NIM : 10111473

**NAMA : CITRAWATI ISRA
SALEKHAH**

**PEMBIMBING : EDNAWATI RAINARLI, S. SI,
M. SI**

LATAR BELAKANG

- Penelitian sebelumnya menggunakan metode K-Nearest Neighbor dengan nilai akurasi mencapai 60% pada $k=5$.
- Hasil penelitian oleh Fatimah Wulandini menunjukkan bahwa metode Support Vector Machine memiliki tingkat akurasi yang paling baik dalam mengklasifikasikan informasi teks Bahasa Indonesia pada situs web dibandingkan dengan metode K-Nearest Neighbor, Naïve Bayes Classifier, dan C.4.5 Decision Tree dengan akurasi mencapai 92.5%.
- Pada penelitian yang dilakukan Jason D. M. Renie juga menunjukkan bahwa metode Support Vector Machine menghasilkan performansi yang lebih baik dalam menyelesaikan klasifikasi teks multikelas dibandingkan metode Naïve Bayes Classifier.

MAKSUD DAN TUJUAN

MAKSUD

Mengimplementasikan metode Multi Class Support Vector Machine untuk klasifikasi emosi pada lirik lagu Bahasa Indonesia.

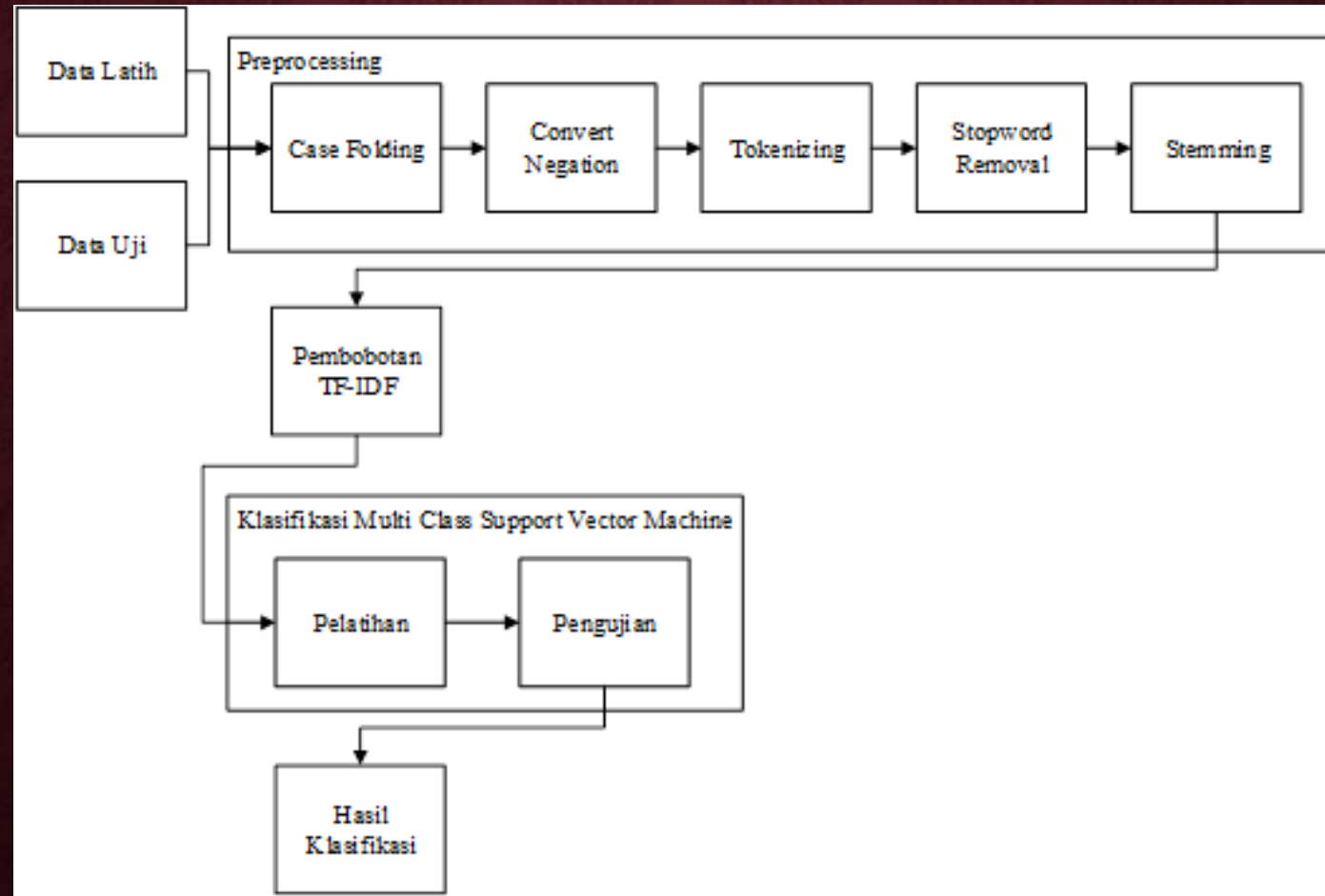
TUJUAN

Mengetahui kinerja dari penggunaan metode Multi Class Support Vector Machine dalam melakukan klasifikasi emosi pada lirik lagu Bahasa Indonesia.

BATASAN MASALAH

- Data latih yang digunakan adalah data International Survey On Emotion Antecedents And Reaction (ISEAR).
- Data uji yang digunakan adalah lirik lagu Bahasa Indonesia dengan format (.txt).
- Kategori emosi yang digunakan adalah senang, sedih, marah, takut, dan bersalah.
- Tahapan preprocessing meliputi case folding, convert negation, tokenizing, stopword removal dan stemming.
- Algoritma stemming adalah algoritma Porter Stemmer Bahasa Indonesia.
- Fungsi kernel yang digunakan adalah Gaussian Radian Basic Function

ANALISIS PROSES



PREPROCESSING

Case Folding

Pemrosesan teks dimana semua huruf diubah menjadi huruf kecil dan menghilangkan karakter selain a-z.

Contoh:

Ketika cinta pertama saya menelepon saya dan mengatakan kepada saya bahwa ia akan selalu peduli kepada saya tidak peduli apa yang saya lakukan.



ketika cinta pertama saya menelepon saya dan mengatakan kepada saya bahwa ia akan selalu peduli kepada saya tidak peduli apa yang saya lakukan

PREPROCESSING

Convert Negation

Pemrosesan teks dimana semua dokumen yang mengandung kata-kata negasi digabungkan dengan kata setelah kata negasi tersebut.

Contoh:

ketika cinta pertama saya menelepon saya dan mengatakan kepada saya bahwa ia akan selalu peduli kepada saya **tidak peduli** apa yang saya lakukan



ketika cinta pertama saya menelepon saya dan mengatakan kepada saya bahwa ia akan selalu peduli kepada saya **tidakpeduli** apa yang saya lakukan

PREPROCESSING

Tokenizing

Pemrosesan teks dimana dilakukan pemotongan kalimat berdasarkan tiap kata yang menyusunnya.

Contoh:

ketika cinta pertama saya menelepon saya dan mengatakan kepada saya bahwa ia akan selalu peduli kepada saya tidakpeduli apa yang saya lakukan



- | | | | |
|-------------|-----------------|---------------|-----------|
| • ketika | • dan | • akan | • apa |
| • cinta | • mengatak
n | • selalu | • yang |
| • pertama | • kepada | • peduli | • saya |
| • saya | • saya | • kepada | • lakukan |
| • menelepon | • bahwa | • saya | |
| • saya | • ia | • tidakpeduli | |

PREPROCESSING

Stopword Removal

Pemrosesan teks dimana menghilangkan kata-kata yang termasuk dalam daftar atau kamus stopwords

Contoh:

- | | | | |
|-------------|--------------|---------------|-----------|
| • ketika | • dan | • akan | • apa |
| • cinta | • mengatakan | • selalu | • yang |
| • pertama | • kepada | • peduli | • saya |
| • saya | • saya | • kepada | • lakukan |
| • menelepon | • bahwa | • saya | |
| • saya | • ia | • tidakpeduli | |



- | | |
|--------------|---------------|
| • cinta | • peduli |
| • menelepon | • tidakpeduli |
| • mengatakan | • lakukan |

PREPROCESSING

Stemming

Pemrosesan teks dimana mengubah kata yang memiliki imbuhan ke dalam bentuk dasarnya.

Contoh:

- cinta
- menelepon
- mengatakan
- peduli
- tidakpeduli
- lakukan



- cinta
- telepon
- kata
- peduli
- tidakpeduli
- laku

PEMBOBOTAN TF-IDF

- Hitung tf (term frequency)
- Hitung df (document frequency)
- Hitung idf = $\log (D/df)$
- Hitung $w = tf * idf$

Dimana

tf = banyaknya kata yang dicari

Df = banyaknya dokumen yang mengandung tiap kata

D = banyaknya data kalimat/dokumen

W = weight atau bobot tiap kata

PEMBOBOTAN TF-IDF

Pernyataan	Kata
P1	cinta telepon kata peduli tidakpeduli laku
P2	sedih dengar pacar dua teman baik
P3	guru sekolah salah kelas tidakmematuhi bicara tidakmelakukan kerja sore jalan salah marah
P4	takut muda sisa rumah
P5	salah sadar timbang materi rawat kerabat egois

Kata	TF					df	idf	TF-IDF				
	P1	P2	P3	P4	P5			P1	P2	P3	P4	P5
cinta	1	0	0	0	0	1	0.699	0.699	0	0	0	0
telepon	1	0	0	0	0	1	0.699	0.699	0	0	0	0
kata	1	0	0	0	0	1	0.699	0.699	0	0	0	0
peduli	1	0	0	0	0	1	0.699	0.699	0	0	0	0
tidakpeduli	1	0	0	0	0	1	0.699	0.699	0	0	0	0
laku	1	0	0	0	0	1	0.699	0.699	0	0	0	0
sedih	0	1	0	0	0	1	0.699	0	0.699	0	0	0
dengar	0	1	0	0	0	1	0.699	0	0.699	0	0	0
pacar	0	1	0	0	0	1	0.699	0	0.699	0	0	0
dua	0	1	0	0	0	1	0.699	0	0.699	0	0	0
teman	0	1	0	0	0	1	0.699	0	0.699	0	0	0
baik	0	1	0	0	0	1	0.699	0	0.699	0	0	0
guru	0	0	1	0	0	1	0.699	0	0	0.699	0	0
sekolah	0	0	1	0	0	1	0.699	0	0	0.699	0	0
salah	0	0	2	0	1	3	0.222	0	0	0.444	0	0.222
kelas	0	0	1	0	0	1	0.699	0	0	0.699	0	0

KLASIFIKASI MULTI CLASS SUPPORT VECTOR MACHINE

PELATIHAN

1. Ubah bobot tf-idf data latih ke dalam bentuk SVM.
2. Memetakan data latih menggunakan kernel Gaussian.

$$K(x, y) = \exp\left(\frac{-|x-y|^2}{2\sigma^2}\right)$$

3. Menentukan sejumlah support vector dengan cara menghitung nilai $\alpha_1, \dots, \alpha_N$ (N=sejumlah data pelatihan) menggunakan quadratic programming

$$Q(a) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j x_i x_j \text{ dengan syarat } \alpha_i > 0 \text{ (} i = 1, 2, \dots, l \text{) dan } \sum_{i=1}^l \alpha_i y_i = 0$$

4. Solusi bidang pemisah didapatkan dengan rumus $w = \sum \alpha_i y_i x_i$; $b = y_k - w^T x_k$ untuk setiap x_k , dengan $\alpha_k \neq 0$

KLASIFIKASI MULTI CLASS SUPPORT VECTOR MACHINE

PENGUJIAN

1. Ubah bobot tf-idf data uji ke dalam bentuk SVM.
2. Memetakan data uji menggunakan kernel Gaussian.

$$K(x, y) = \exp\left(\frac{-|x-y|^2}{2\sigma^2}\right)$$

3. Menghitung fungsi keputusan: $f_i = K(x_i, x_d)w_i + b_i$

Dimana: $i = 1$ sampai k ; x_i = support vector ; x_d = data pengujian

4. Menentukan nilai f_i yang paling maksimal. Kelas i dengan f_i terbesar adalah kelas dari data pengujian.