

SUPPORT VECTOR MACHINE DAN INFORMATION GAIN UNTUK KLASIFIKASI EMOSI PADA LIRIK LAGU

Eva Nur Azizah¹, Ednawati Rainarli²

^{1,2} Universitas Komputer Indonesia

Jl. Dipati Ukur No.112-116, Lebakgede, Cobleng, Kota Bandung, Jawa Barat 40132

E-mail : eva@email.unikom.ac.id¹, ednawati.rainarli@email.unikom.ac.id²

ABSTRAK

Salah satu bentuk teks yang dapat mengekspresikan emosi adalah lirik lagu. Penelitian sebelumnya yang dilakukan oleh Citrawati Isra Salekhah yang melakukan penelitian terhadap lirik lagu didapatkan akurasi hanya 36,66%. Hal ini dikarenakan data latih yang digunakan bahasa Inggris sehingga saat diterjemahkan kedalam bahasa Indonesia terjadi kesalahan makna kosakata (leksikal) yang mengakibatkan kesalahan makna yang disampaikan. Oleh karena itu dalam penelitian ini akan digunakan data latih lagu bahasa Indonesia yang telah divalidasi oleh ahli bahasa. Seleksi fitur *Information Gain* akan ditambahkan dengan tujuan untuk mendapatkan fitur yang relevan dalam mendeteksi emosi pada lagu. Hasil pengujian menunjukkan bahwa penggunaan *Support Vector Machine* dan seleksi fitur *Information Gain* dengan data latih sebanyak 1000 baris lirik dan data uji sebanyak 300 baris lirik menghasilkan akurasi terbaik sebesar 73.3 %. Hasil ini diperoleh dengan menggunakan fungsi kernel Polynomial derajat 1. Sedangkan untuk hasil pengujian pada emosi sebanyak 20 lagu menunjukkan bahwa penggunaan *Support Vector Machine* dan seleksi fitur *Information Gain* mendapatkan hasil akurasi terbaik sebesar 70%. Hal ini menunjukkan bahwa penggunaan seleksi fitur *Information Gain* bisa meningkatkan akurasi karena fitur yang tidak relevan terhadap target klasifikasi telah berkurang.

Kata kunci : *Support Vector Machine*, *Information Gain*, klasifikasi, emosi, lagu

1. PENDAHULUAN

Emosi dapat diungkapkan oleh seseorang dari ucapan, ekspresi wajah, nada suara dan teks [1]. Salah satu bentuk teks yang dapat mengekspresikan emosi adalah lirik lagu. Lirik lagu merupakan karya sastra yang berupa teks yang isinya dapat mengungkapkan curahan perasaan pribadi, pikiran dan emosi dari penulis lagu. Oleh karena itu lirik lagu dapat digunakan sebagai objek penelitian dalam klasifikasi emosi.

Pada penelitian sebelumnya tentang klasifikasi berita yang dilakukan oleh Siti Nur Asiyah menunjukkan bahwa metode *Support Vector Machine* memiliki kinerja lebih baik dibandingkan dengan metode K-NN, dimana akurasi di *Support Vector Machine* mencapai 93,2% [2]. Akan tetapi hasil penelitian Siti N berbeda dengan yang didapatkan oleh Citrawati Isra Salekhah yang melakukan penelitian terhadap lirik lagu bahasa Indonesia menggunakan metode *Multi Class Support Vector Machine* dengan pembobotan TF-IDF. Hasilnya didapatkan bahwa nilai akurasi dari penelitian tersebut hanya 36,66%. Hasil pengujian penelitian tersebut menunjukkan pada pelatihan dengan menggunakan data yang lebih banyak, akurasinya cenderung lebih kecil akurasinya dari data yang lebih sedikit. Hal ini dikarenakan data latih yang digunakan bahasa Inggris sehingga saat diterjemahkan kedalam bahasa Indonesia terjadi kesalahan makna kosakata (leksikal) yang mengakibatkan kesalahan makna yang disampaikan [3]. Oleh karena itu dalam penelitian ini akan digunakan data latih lagu bahasa Indonesia yang telah divalidasi oleh ahli bahasa. Seleksi fitur *Information Gain* akan ditambahkan dengan tujuan untuk mendapatkan fitur yang relevan dalam mendeteksi emosi pada lagu.

Support Vector Machine merupakan metode yang banyak digunakan terutama pada proses klasifikasi data. Selain itu, SVM juga mempunyai keunggulan dalam mengolah data dengan dimensi tinggi tanpa mengurangi performa [4]. Hal ini dapat ditunjukkan pada penelitian yang dilakukan oleh Joachims yang membandingkan SVM untuk text categorization dengan beberapa metode lain yaitu *Bayes*, *Rocchio*, R4.5, dan K-NN diketahui bahwa SVM menghasilkan performa yang baik, mengungguli metode lainnya secara substansial dan signifikan [4]. Penggunaan metode *Support Vector Machine* menunjukkan hasil yang lebih baik dari segi ketepatan maupun kecepatan pembuatan model [12].

Information Gain merupakan salah satu seleksi fitur yang dapat digunakan untuk memilih fitur terbaik dan bersifat informatif. *Information Gain* dapat melihat setiap fitur untuk memprediksi label kelas yang benar karena memilih nilai tertinggi dan bersifat penting sehingga dapat meningkatkan kinerja algoritma klasifikasi [5]. Hal ini ditunjukkan pada penelitian yang dilakukan Bagus Setya dalam

kategorisasi teks dengan metode *Support Vector Machine* bahwa penambahan seleksi fitur *Information Gain* dapat menunjukkan peningkatan precision sebesar 15% dan *recall* sebesar 13% dibandingkan dengan seleksi fitur *Particle Swarm Optimization* (PSO) [6].

Oleh karena itu pada penelitian ini akan digunakan metode *Support Vector Machine* dan seleksi fitur *Information Gain* untuk melihat akurasi pada metode tersebut.

Maksud dari penelitian ini adalah mengimplementasikan seleksi fitur *Information Gain* pada metode klasifikasi *Support Vector Machine* untuk klasifikasi emosi pada lirik lagu. Sedangkan tujuan yang ingin dicapai pada penelitian ini adalah mengukur kinerja algoritma seleksi fitur *Information Gain* pada algoritma klasifikasi metode *Support Vector Machine* untuk menentukan emosi pada lirik lagu.

2. LANDASAN TEORI

Emosi

Emosi adalah gambaran keadaan seseorang yang disebabkan oleh kejadian penting. Emosi meliputi keadaan mental sadar, gangguan jasmani pada beberapa organ tubuh, dan pengenalan ekspresi pada wajah [7].

Jenis emosi dasar pada manusia ada 5 macam yaitu :

- Senang, menggambarkan seseorang yang berhasil atau bergerak menuju keberhasilan suatu tujuan
- Sedih, menggambarkan kegagalan atau kerugian
- Marah, menggambarkan seseorang yang frustrasi dari peran atau tujuan yang dirasakan orang lain
- Takut, menggambarkan ancaman fisik atau sosial untuk diri sendiri

Preprocessing

Preprocessing adalah tahapan untuk mempersiapkan teks menjadi data yang akan diolah di tahapan berikutnya. Teks yang akan dilakukan pada proses ini terdapat *noise* dan terdapat struktur yang tidak baik [8]. Tahapan *preprocessing* dalam penelitian ini yaitu *case folding*, *convert negation*, *filtering*, *tokenizing* dan *stopword removal*.

a. Case folding

Case folding merupakan proses yang dilakukan untuk menyeragamkan karakter pada data (dokumen/teks). Pada tahap ini semua karakter diseragamkan menjadi huruf kecil (*lowercase*) [8].

b. Convert Negation

Convert Negation merupakan proses konversi kata-kata negasi yang terdapat pada suatu kalimat, karena kata negasi mempunyai pengaruh dalam merubah nilai emosi pada sebuah kalimat [11]. Contoh kata negasi dalam bahasa Indonesia adalah “tak”, “tidak”, dan “gak”.

c. Tokenizing

Tokenizing yaitu penguraian lirik lagu yang berupa kalimat – kalimat menjadi kata-kata . Proses *tokenizing* dilakukan dengan cara memisahkan kata berdasarkan spasi yang terletak diantara dua kata [11].

d. Filtering

Filtering adalah tahap pemilihan kata – kata penting, yaitu kata apa saja yang akan mewakili isi dari dokumen. Pada tahap ini karakter selain huruf dihilangkan dan dianggap delimiter [11].

e. Stopword Removal

Stopword Removal adalah proses untuk menghilangkan kata yang tidak relevan pada dokumen teks dengan cara membandingkan dengan *stoplist* yang ada. *Stoplist* berisi sekumpulan kata yang tidak relevan namun sering muncul dalam sebuah dokumen. Pada penelitian ini *stoplist* yang digunakan diambil dari penelitian Fadillah Z. Tala [9].

Pembobotan TF-IDF

Term Frequency – Inverse Document Frequency digunakan untuk menentukan nilai bobot sebuah kata di dalam banyaknya dokumen [16]. Pembobotan dapat diperoleh berdasarkan jumlah kemunculan suatu kata (*term*) dalam koleksi dokumen *term frequency* (*tf*) dan jumlah kemunculan term dalam koleksi dokumen *inverse document frequency* (*idf*). Nilai *idf_t* sebuah term dapat dihitung menggunakan persamaan (1):

$$idf_t = \log\left(\frac{N}{df_t}\right) \quad (1)$$

Keterangan :

N = jumlah dokumen

df = kemunculan dokumen yang mengandung *term*

Adapun persamaan yang digunakan untuk menghitung bobot (*W*) masing – masing dokumen yaitu menggunakan persamaan (2) :

$$W_t = tf_{dt} * idf_t \quad (2)$$

Keterangan :

tf_{dt} = jumlah kemunculan *term t* terhadap dokumen *d*

idf_t = hasil dari *inverse document frequency* dari *term t*

Information Gain

Information Gain adalah salah satu algoritma seleksi fitur yang digunakan untuk memilih fitur terbaik. Nilai *Information Gain* yang didapat akan digunakan untuk menyeleksi fitur menggunakan *threshold* sehingga menghasilkan fitur terbaik [15]. Nilai *Information Gain* dapat didefinisikan dengan persamaan (3) [10] :

$$IG(t) = -\frac{A+C}{N} \log\left(\frac{A+C}{N}\right) + \frac{A}{N} \log\left(\frac{A}{A+B}\right) + \frac{C}{N} \log\left(\frac{C}{C+D}\right) \quad (3)$$

Keterangan :

A = jumlah dokumen dikelas *k* yang mengandung term *t*

B = jumlah dokumen diluar dikelas *k* yang mengandung term *t*

C = jumlah dokumen dikelas k yang tidak mengandung term t
 D = jumlah dokumen diluar dikelas k yang tidak mengandung term t
 N = total dokumen

Setelah nilai *Information Gain* didapat, maka dilakukan langkah selanjutnya adalah menentukan nilai *threshold*. Nilai *Information Gain* yang paling kecil dapat dijadikan sebagai nilai *threshold* dengan akurasi terbaik [13].

Support Vector Machine

Support Vector Machine adalah algoritma *supervised learning* yang digunakan untuk analisis klasifikasi. *Support Vector Machine* merupakan *binary classifier* yang membagi data menjadi dua kelas yang disebut *hyperlane* [14]. *Hyperlane* ini tepat berada di tengah – tengah kedua kelas. Namun *Support Vector Machine* telah dikembangkan agar dapat bekerja dalam kasus *non linear* dengan menggunakan konsep kernel pada ruang dimensi tinggi. Fungsi kernel digunakan untuk memetakan set fitur awal yang lebih rendah pada set fitur baru yang lebih tinggi. Macam – macam fungsi kernel dapat dilihat pada Tabel 1 [15]:

Tabel 1. Fungsi Kernel

Nama Kernel	Definisi Fungsi
Linear	$K(x, x_k) = x \cdot x_k^T$
Polynomial	$K(x, x_k) = (x \cdot x_k^T + 1)^d$
Gaussian RBF	$K(x, x_k) = \exp\left(\frac{-\ x - x_k\ ^2}{2 \cdot \sigma^2}\right)$
Sigmoid	$K(x, x_k) = \tanh[K x_k^T \cdot x + \theta]$

Hyperlane klasifikasi SVM dapat dinyatakan dengan persamaan (4) :

$$w \cdot x_i + b = 0 \quad (4)$$

Jika data adalah x_i dan label data adalah $y_i = -1$ (negatif) maka dapat dinyatakan dengan persamaan (5) sebagai berikut :

$$w \cdot x_i + b \leq -1 \quad (5)$$

Sedangkan jika data adalah x_i dan label data adalah $y_i = +1$ (positif) maka dapat dinyatakan dengan persamaan (6) :

$$w \cdot x_i + b \geq +1 \quad (6)$$

Untuk mencari *hyperlane* terbaik dapat menggunakan *Quadratic Programming* (QP) yaitu dengan meminimalkan persamaan $\frac{1}{2} \|w\|^2$ menggunakan fungsi *Lagrange Multiplier* yang dijabarkan sebagai menjadi persamaan (7):

$$L_p = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i y_i (w \cdot x_i - b - \sum_{i=1}^N \alpha_i y_i + \sum_{i=1}^N \alpha_i) \quad (7)$$

Dimana α_i adalah *lagrange multiplier* yang berkorespondensi dengan x_i . Nilai α_i merupakan nilai nol atau positif ($\alpha_i \geq 0$). Persamaan diatas akan berubah menjadi dualitas *Lagrange multiplier* berupa L_D dengan memaksimalkan persamaan (8) :

Maksimalkan :

$$L_D = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^N \alpha_i \alpha_j y_i y_j x_i x_j \quad (8)$$

Dengan syarat 1:

$$\sum_{i=1}^N \alpha_i y_i = 0 \quad (9)$$

Dan syarat 2:

$$\alpha_i \geq 0, i = 1, 2, \dots, N \quad (10)$$

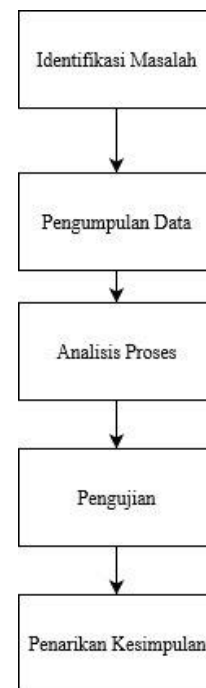
Dengan $x_i x_j$ merupakan *dot-product* dua data dalam data latih. Setelah solusi *Quadratic Programming* ditemukan (nilai α_i), maka kelas dari data uji dapat ditentukan dengan persamaan (11) :

$$f(\Phi(x)) = \text{sign}(w \cdot \Phi(z) + b) = \text{sign}(\sum_{i=1}^N \alpha_i y_i \Phi(x_i) \cdot \Phi(z) + b) \quad (11)$$

Dimana N adalah jumlah data yang menjadi *support vector*, x_i adalah *support vector*, dan z adalah data uji yang akan di prediksi.

3. METODE PENELITIAN

Tahapan yang dilakukan dalam penelitian ini terdapat pada Gambar 1. sebagai berikut :



Gambar 1. Tahapan penelitian yang digunakan

Berikut ini merupakan penjelasan dari setiap tahapan yang dilakukan pada penelitian ini:

a. Identifikasi Masalah

Pada tahapan identifikasi masalah dilakukan analisis mengenai masalah yang dihadapi pada penelitian yang dilakukan.

b. Pengumpulan Data

Pada tahapan pengumpulan data yaitu mengumpulkan data literatur dan data lirik lagu. Literatur yang dibutuhkan adalah mengenai metode *Support Vector Machine* dan *Information Gain*. Sedangkan data lirik lagu diambil dari berbagai situs <https://liriklaguindonesia.net/>, <https://no1lyrics.com/>, dan situs lainnya. Setiap baris lirik akan divalidasi oleh ahli bahasa berdasarkan jenis emosi.

c. Analisis Proses

Pada tahap analisis proses dilakukan analisis mengenai proses klasifikasi emosi. Analisis yang dilakukan meliputi analisis masalah, analisis *preprocessing*, analisis seleksi fitur *Information Gain* dan analisis metode *Support Vector Machine*.

d. Pengujian

Pada tahapan ini akan dilakukan pengujian terhadap hasil klasifikasi dengan menggunakan *Confusion Matrix*.

e. Penarikan Kesimpulan

Pada tahapan ini yaitu menentukan kesimpulan berdasarkan hasil pengujian yang telah dilakukan..

4. HASIL DAN PEMBAHASAN

Analisis Masalah

Masalah yang ditemukan adalah pada penelitian sebelumnya yaitu data latih yang digunakan berbahasa inggris sehingga saat diterjemahkan kedalam bahasa indonesia terjadi kesalahan makna kosakata. Solusi yang diberikan adalah mengganti data latih dengan lirik lagu bahasa Indonesia yang telah divalidasi oleh ahli bahasa dan menambahkan seleksi fitur *Information Gain* karena seleksi fitur dapat mendapatkan fitur yang relevan dalam mendeteksi emosi pada lagu sehingga dapat meningkatkan kinerja algoritma klasifikasi.

Analisis Data Masukan

Data akan dibagi menjadi dua jenis yaitu data latih dan data uji. Data latih yang digunakan adalah kumpulan kalimat sebanyak 1000 baris lirik lagu yang sudah dikelompokkan berdasarkan jenis emosi yang ditentukan. Sedangkan data uji yang digunakan adalah kumpulan kalimat sebanyak 300 baris lirik lagu. Dalam penelitian ini ada 4 kategori emosi yang sudah ditentukan yaitu senang, sedih, marah dan takut.

Analisis Proses

Tahapan yang digunakan untuk mengetahui jenis emosi dalam lirik lagu pada penelitian ini dibagi menjadi dua tahap, yaitu tahap pelatihan dan tahap pengujian.

Dalam tahap pelatihan data latih akan melalui empat proses utama yaitu proses *preprocessing*, pembobotan *TF-IDF*, seleksi fitur *Information Gain* dan proses pelatihan *Support Vector Machine*. Dalam tahap pengujian, data uji yang dimasukkan melalui proses yang sama seperti data latih yaitu proses *preprocessing*, pembobotan *TF-IDF*, lalu proses pengujian dengan *Support Vector Machine*. Hasil dari tahap pengujian ini adalah hasil klasifikasi berupa kategori emosi dari data uji yang telah dimasukkan. Lalu untuk pengujian sistem disajikan dalam *confuss matrix*. Hasil dari pengujian tersebut adalah perbandingan akurasi dari metode *Support Vector Machine* tanpa seleksi fitur *Information Gain* dan *Support Vector Machine* dengan seleksi fitur *Information Gain*.

Pengujian Akurasi

Pengujian akurasi dibagi menjadi tiga bagian, dimana setiap bagian tersebut akan digunakan jumlah data uji yang bervariasi. Pada pengujian akurasi pertama yaitu untuk mengetahui model klasifikasi, data uji yang digunakan sama dengan data latih yang digunakan yaitu sebanyak 1000 baris lirik. Pada pengujian kedua digunakan data latih sebanyak 1000 baris lirik dan data uji sebanyak 300 baris lirik. Pada pengujian ketiga yaitu pengujian emosi lagu dengan menguji sebanyak 20 lagu dan data latih sebanyak 1000 baris lirik.

Hasil pengujian akurasi berisi hasil pengujian dari akurasi menggunakan metode *Support Vector Machine* dan *Information Gain*. Berikut hasil pengujiannya:

a. Hasil Pengujian Model Klasifikasi

Berikut merupakan model hasil kerja klasifikasi dengan menggunakan data latih sama dengan data uji dengan jumlah data 1000 baris lirik, hasil akurasinya dapat dilihat pada Tabel 2 :

Tabel 2. Hasil Pengujian Model Klasifikasi

	SVM	SVM+IG Threshold= 0.00181	SVM+IG Threshold= 0.00183
Linear	88,8%	86,7%	80,4%
RBF γ=1	94,6%	93,2%	84,2%
RBF γ=2	96,2%	95,3%	85,1%
RBF γ=3	96,2%	95,3%	85,1%
Polynom n=1	94,5%	93,4%	85 %
Polynom n=2	96,2%	95,3%	85,1%
Polynom n=3	96%	95,3%	85,1%

Hasil pengujian tersebut menunjukkan bahwa model terbaik dihasilkan oleh *Support Vector Machine* dengan fungsi kernel RBF dan Polynomial menghasilkan akurasi sebesar 96,2%. Hal ini menunjukkan bahwa penggunaan seleksi fitur *Information Gain* tidak terlalu berpengaruh untuk data yang sama.

b. Hasil Pengujian dengan Data uji yang berbeda

Berikut merupakan hasil kerja klasifikasi dengan menggunakan data latih sebanyak 1000 baris lirik dan data uji sebanyak 300 baris lirik, yang dapat dilihat pada Tabel 3 :

Tabel 3. Hasil Pengujian dengan 300 baris lirik

	SVM	SVM+I Threshold= 0.00181	SVM+IG Threshold= 0.00183
Linear	64,6%	66,6%	63,3%
RBF $\gamma=1$	66%	68%	65%
RBF $\gamma=2$	67,6%	70%	65,3%
RBF $\gamma=3$	65%	67%	66,3%
Polynom n=1	70,3%	73,3%	65%
Polynom n=2	65%	66,3%	62%
Polynom n=3	55,6%	61,3%	62%

Hasil pengujian tersebut menunjukan bahwa metode terbaik dihasilkan oleh *Support Vector Machine* dengan fungsi kernel Polynom derajat 1 menggunakan seleksi fitur *Information Gain*. Hasil akurasi yang diperoleh sebesar 73,3%. Hal ini menunjukan bahwa penggunaan seleksi fitur *Information Gain* bisa meningkatkan akurasi karena fitur yang tidak relevan terhadap target klasifikasi telah berkurang.

c. Pengujian Emosi Lagu

Pada pengujian ini dibagi menjadi 2 yaitu pengujian menggunakan *Support Vector Machine* tanpa seleksi fitur *Information Gain* dan *Support Vector Machine* dengan seleksi fitur *Information Gain*. Fungsi kernel yang digunakan yaitu fungsi kernel polynom derajat 1.

Hasil pengujian matriks konfusi untuk pengujian lagu tanpa seleksi fitur dapat dilihat pada Tabel 4 :

Tabel 4. Matriks Konfusi Pengujian Lagu tanpa seleksi fitur

F_{ij}		Kelas prediksi(j)			
		Kelas 1	Kelas 2	Kelas 3	Kelas 4
Kelas asli (i)	Kelas 1	5	0	3	0
	Kelas 2	3	2	0	1
	Kelas 3	0	1	0	0
	Kelas 4	1	0	0	4

Hasil pengujian tersebut menunjukan bahwa jumlah lagu yang diklasifikasi secara benar sebanyak 11 lagu dan hasil yang salah sebanyak 9 lagu. Hasil menunjukan bahwa total lagu dikelas 1 (senang) sebanyak 8 lagu dengan 5 lagu terprediksi secara benar di kelas senang dan 3 lagu terprediksi di kelas 3 (marah).

Maka perhitungan nilai akurasi dengan melihat matriks konfusi didapat sebesar 55%.

$$Akurasi = \frac{11}{20} \times 100\% = 55\%$$

Sedangkan hasil pengujian matriks konfusi untuk pengujian lagu menggunakan seleksi fitur dapat dilihat pada Tabel 5 :

Tabel 5. Matriks Konfusi Pengujian Lagu dengan seleksi fitur

F_{ij}		Kelas prediksi(j)			
		Kelas 1	Kelas 2	Kelas 3	Kelas 4
Kelas asli (i)	Kelas 1	8	0	0	0
	Kelas 2	3	2	0	1
	Kelas 3	0	0	1	0
	Kelas 4	2	0	0	3

Hasil pengujian tersebut menunjukan bahwa jumlah lagu yang diklasifikasi secara benar sebanyak 14 lagu dan hasil yang salah sebanyak 6 lagu. Hasil menunjukan bahwa total lagu dikelas 2 (sedih) sebanyak 6 lagu dengan 2 lagu terprediksi secara benar di kelas sedih, 3 lagu terprediksi dikelas senang dan 1 lagu di kelas 4 (takut).

Maka perhitungan nilai akurasi dengan melihat matriks konfusi didapat sebesar 70%.

$$Akurasi = \frac{14}{20} \times 100\% = 70\%$$

5. KESIMPULAN

Hasil pengujian menunjukan bahwa penggunaan *Support Vector Machine* dan seleksi fitur *Information Gain* dengan data latih sebanyak 1000 baris lirik dan data uji sebanyak 300 baris lirik menghasilkan akurasi terbaik sebesar 73.3 %. Hasil ini diperoleh dengan menggunakan fungsi kernel Polynomial derajat 1. Sedangkan untuk hasil pengujian pada emosi sebanyak 20 lagu menunjukan bahwa penggunaan *Support Vector Machine* dan seleksi fitur *Information Gain* mendapatkan hasil akurasi terbaik sebesar 70%. Hal ini menunjukan bahwa penggunaan seleksi fitur *Information Gain* bisa meningkatkan akurasi karena fitur yang tidak relevan terhadap target klasifikasi telah berkurang. Saran untuk kedepannya bisa menerapkan *deep learning* untuk pengklasifikasian.

DAFTAR PUSTAKA

- [1] V. V. Ramalingam, A. Pandian, A. Jaiswal, and N. Bhatia, "Emotion detection from text," *J. Phys. Conf. Ser.*, vol. 1000, no. 1, 2018.
- [2] S. N. Asiyah and K. Fithriasari, "Klasifikasi Berita Online Menggunakan Metode Support Vector Machine dan K- Nearest Neighbor," *J. Sains dan Seni ITS*, vol. 5, no. 2, pp. 317–322, 2016.
- [3] R. Baharuddin, "Kesalahan Makna Leksikal pada Terjemahan Teks Bahasa Indonesia ke dalam Bahasa Inggris," *Dialekt. J. Lang. Lit. Math. Educ.*, vol. 1, no. 1, pp. 42–55, 2015.

- [4] I. K. Purnamawan, "Support Vector Machine Pada Information Retrieval," *Jptk*, vol. 12, no. 2, pp. 173–180, 2015.
- [5] I. Maulida, A. Suyatno, and H. R. Hatta, "Seleksi Fitur Pada Dokumen Abstrak Teks Bahasa Indonesia Menggunakan Metode Information Gain," *Jsm Stmik Mikroskil*, vol. 17, no. 2, pp. 249–258, 2016.
- [6] B. S. Rintyarna and A. Z. Arifin, "Seleksi Fitur Dua Tahap Menggunakan Information Gain dan Artificial Bee Colony untuk Kategorisasi Teks Berbasis Support Vector Machine," *Systemic*, vol. 1, no. 2, pp. 22–26, 2015.
- [7] S. Sumpeno, "Klasifikasi Emosi Untuk Teks Bahasa Indonesia," *Semin. Nas. Pascasarj. IX – ITS*, 2009.
- [8] I. Feinerer, "Mining Text Data," *R News*, vol. 8, pp. 51–88, 2012.
- [9] F. Z. Tala, "A Study Of Stemming Effects On Information Retrieval In Bahasa Indonesia," *M.Sc. Thesis, Append. D, Vol. Pp, Pp. 39–46*, 2003.
- [10] M. Fatih And S. Bayir, "Examining The Impact Of Feature Selection Methods On Text Classification," *IJACSA) Int. J. Adv. Comput. Sci. Appl., Vol. 8, No. 12, Pp. 380–388*, 2017
- [11] R. Fahreza Nur Firmansyah, M. Fauzi, T. Afirianto, "Sentiment Analysis Pada Review Aplikasi Mobile Menggunakan Metode Naïve Bayes Dan Query Expansion," *Vol. 8, 2016*.
- [12] E. Rainarli and A. Romadhan, "Perbandingan Simple Logistic Classifier dengan Support Vector Machine dalam Memprediksi Kemenangan Atlet," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 3, no. 2, pp. 87–91, 2017.
- [13] R. M. Alfajri, Y. H. Chrisnanto, And R. Yuniarti, "Pengklasifikasian Kemampuan Akademik Mahasiswa Menggunakan Metode Information Gain Dan Naive Bayes Classifier Dalam Prediksi Penyelesaian Studi Tepat Waktu," *Pros. Snst, No. 7, pp. 144–149*, 2016.
- [14] E. Prasetyo, *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*, Andi, 2014.
- [15] Suyanto, *Data Mining Untuk Klasifikasi Dan Klasterisasi Data*, Informatika, 2017.
- [16] D. S. Harjanto, S. N. Endah, and N. Bahtiar, "Sistem Temu Kembali Informasi pada Dokumen Teks Menggunakan Metode Term Frequency Inverse Document Frequency (TF-IDF)," *J. Sains dan Mat.*, vol. 20, no. 3, pp. 64–70, 2012.