

# Pengenalan Emosi Dalam Musik Berdasarkan *Musical Features* Menggunakan *Support Vector Regression*

Emilia Widiyanti<sup>\*1)</sup>, Sukmawati Nur Endah, S.Si., M.Kom<sup>\*2)</sup>

<sup>\*</sup>Jurusan Ilmu Komputer / Informatika, Universitas Diponegoro

<sup>1)</sup>emilia.widiyanti@gmail.com, <sup>2)</sup>sukmawati020578@gmail.com

## *Abstrak*

*Musik dibuat untuk menyampaikan emosi dan seringkali dimanfaatkan dalam berbagai kegiatan sehari-hari. Music Emotion Recognition atau pengenalan emosi dalam musik menjadi salah satu bidang penelitian yang ikut berkembang seiring dengan perkembangan jenis dan pemanfaatan musik. Penelitian ini menyajikan hasil pengenalan emosi pada musik dengan musical features menggunakan Support Vector Regression dengan jenis pelatihan  $\epsilon$ -Support Vector Regression dan  $\nu$ -Support Vector Regression serta kombinasi fitur terbaik yang menghasilkan model terbaik. Data yang digunakan sejumlah 165 data musik yang berbentuk musik soundtrack instrumental. Dari penelitian ini dihasilkan dua model terbaik menggunakan pelatihan  $\nu$ -SVR. Model yang dihasilkan yaitu model pengenalan angle dengan masukan fitur terbaik adalah fitur Pitch dan Energy, dan model pengenalan distance dengan masukan fitur terbaik Zero Crossing Rate dan Beat. Model dihasilkan dengan nilai parameter pelatihan model untuk  $cost=2^7$ ,  $gamma=2^{-7}$  dan  $nu=2^{-2}$  pada model angle dan  $cost=2^7$ ,  $gamma=2^{-8}$  dan  $nu=2^{-2}$  pada model distance. Pengenalan dengan kedua model tersebut menghasilkan akurasi sebesar 37,75%.*

**Kata kunci :** *Soundtrack, Music Emotion Recognition, Musical Features, Support Vector Regression*

## *Abstract*

*Music was made to convey emotions and has many uses on daily life. Music Emotion Recognition research field has been growing as the development and application of music becomes vary as well. This research produces the result of emotion recognition in music with musical features using Support Vector Regression with  $\epsilon$ -Support Vector Regression and  $\nu$ -Support Vector Regression training along with the optimal feature combination which produces the best model. The data being used are 165 musics in the form of instrumental soundtrack. This research resulted in two best models using the  $\nu$ -SVR training, which the first is angle recognition model with optimal features input of Pitch and Energy, and the second is distance recognition model with optimal features input of Zero Crossing Rate and Beat. Recognition process using those models yields an outcome of 37,35% accuracy that was obtained with parameter values of  $cost=2^7$ ,  $gamma=2^{-7}$  and  $nu=2^{-2}$  in training angle model and parameter values of  $cost=2^7$ ,  $gamma=2^{-8}$  dan  $nu=2^{-2}$  in training distance model.*

**Keywords :** *Soundtrack, Music Emotion Recognition, Musical Features, Support Vector Regression*

## 1 PENDAHULUAN

Musik memiliki peranan penting dalam sejarah manusia dan hampir semua musik dibuat untuk menyampaikan emosi [1]. Emosi yang terkandung dalam musik dapat berupa emosi yang menggambarkan perasaan komposer atau emosi yang berkaitan dengan perasaan pendengar ketika mendengarkan musik tersebut. Musik pun menjadi hal yang tidak dapat lepas dari kegiatan sehari-hari. Tak jarang manusia memanfaatkan emosi yang terkandung dalam musik untuk berbagai tujuan misalnya sebagai pembangkit semangat, sebagai *service/* layanan untuk meningkatkan kenyamanan bagi pembeli pada saat berbelanja, sebagai pembangkit *mood* tertentu dan lain sebagainya.

Salah satu manfaat musik sebagai pembangkit *mood* dimanfaatkan juga dalam industri film dan *game*. Musik digunakan sebagai elemen untuk meningkatkan *impact* dalam sebuah *scene* pada film. Begitupun pada industri *game* yang memanfaatkan musik sebagai penarik perhatian dan menggambarkan kondisi tertentu berdasarkan skenario *game*. Musik pengiring dalam film atau *game* ini biasa disebut sebagai *background music* atau disingkat sebagai BGM. Istilah lain yang dapat digunakan yaitu *original soundtrack* yang disingkat sebagai OST. Musik yang digunakan sebagai BGM atau OST ini pun bermacam-macam jenis dan *genre*-nya. Selain itu, data BGM/ OST ini pun tersebar luas melalui jaringan internet yang menjadikannya cukup populer di kalangan masyarakat.

Perkembangan jenis dan pemanfaatan dari musik ikut berimbas pada peningkatan minat terhadap penelitian di bidang terkait dengan musik. Salah satu bidang penelitian yang diminati diantaranya pengenalan emosi

dalam musik atau *Music Emotion Recognition* (MER). MER menjadi salah satu elemen dalam pengembangan macam-macam aplikasi seperti aplikasi *music retrieval* dan *mood detection*.

Pada penelitian MER, pengenalan emosi dilakukan dengan melalui beberapa jenis pendekatan emosi. Salah satu pendekatan emosi pada MER adalah pendekatan dimensional. Pendekatan dimensional MER memetakan emosi-emosi pada sebuah bidang dimensional.

Penelitian di bidang psikologi telah banyak membahas pemetaan emosi pada sebuah bidang dimensional. Salah satu model emosi yang diperkenalkan yaitu model Thayer [13]. Thayer menyarankan sebuah model emosi yang sederhana namun tangguh dalam mengorganisasi respon emosi yang berbeda yaitu stres dan energi. Pada model Thayer, emosi direpresentasikan pada bidang dua dimensi dengan menggunakan *valence* sebagai representasi stres dan *arousal* sebagai representasi energi. Dikembangkan pula model emosi lain yang menggabungkan sebelas tipe emosi yang berbeda ke dalam model emosi Thayer [4]. Melalui penggabungan tersebut dihasilkan model emosi dua dimensi yang memetakan area dari sebelas jenis emosi yang berbeda, berdasarkan respon emosi tersebut terhadap faktor *valence* dan *arousal*.

Pada musik, *valence* dan *arousal* dapat diamati melalui *musical features* dari sebuah data musik. *Musical features* atau fitur musikal, dideskripsikan sebagai fitur yang ditekankan pada karakteristik musikal dari sebuah data. *Musical features* sering pula dikaitkan dengan elemen-elemen musik yaitu *melody*, *rhythm*, *harmony timbre*, *form*, *dynamics* dan *texture*. Perbedaan persepsi pendengar terhadap emosi dalam musik dapat pula dipengaruhi oleh elemen akustik dari sebuah musik [16], sehingga dirasa

penting untuk mempertimbangkan penggunaan musical features sebagai fitur masukan pada aplikasi MER. Pemilihan fitur yang tepat akan membantu aplikasi MER untuk dapat mengenali emosi dalam musik dengan lebih baik.

Pada saat penelitian dilakukan, belum ada metode standar untuk mengukur dan menganalisa nilai emosi [4], termasuk didalamnya metode standar pengukuran nilai *valence* dan *arousal*. Salah satu metode pada pendekatan dimensional MER yang dapat digunakan adalah dengan memanfaatkan *Support Vector Regression* (SVR). SVR adalah sebuah pengembangan dari *Support Vector Machine* (SVM) yang digunakan untuk menemukan sebuah *mapping function* berupa fungsi regresi dari suatu input dan output. Terdapat dua tipe pelatihan pada SVR yaitu  $\epsilon$ - (dibaca 'epsilon')SVR dan  $\nu$ - (dibaca 'nu')SVR. Jenis pelatihan  $\epsilon$ -SVR menggunakan parameter yang disebut epsilon yang berfungsi sebagai batas error yang diijinkan dalam pelatihan. Berbeda dengan jenis pelatihan  $\nu$ -SVR yang menggunakan parameter yang disebut  $\nu$  (atau  $\nu$ ) sebagai *control parameter* yang mengatur jumlah *support vector* dan *training errors* [16].

SVR digunakan dalam melakukan prediksi nilai parameter pengenalan emosi berdasarkan *groundtruth* data musik yang digunakan dalam pengenalan emosi. *Groundtruth* data musik dibentuk dengan mengaplikasikan sebuah *separation policy* pada *emotion plane* dan dapat dibentuk menggunakan sistem koordinat kartesius maupun polar. Dalam penelitian sebelumnya dipaparkan bahwa klasifikasi emosi menggunakan  $\nu$ -SVR memiliki nilai keakuratan lebih tinggi dibanding dengan SVM dan *Gaussian Mixture Model* yakni mencapai angka 94,55% dengan

pembentukan *groundtruth* yang dilakukan berdasarkan sistem koordinat polar [4].

Pada penelitian ini akan dilakukan pengenalan emosi dalam musik dengan *Musical Features* sebagai fitur masukan menggunakan  $\epsilon$ -SVR dan  $\nu$ -SVR sebagai metode pembelajaran. Pembentukan *groundtruth* berdasarkan sistem koordinat polar dengan data *sountrack* digunakan sebagai data acuan dalam penelitian.

Penelitian ini bertujuan untuk mengenali emosi dalam musik dengan *Musical Features* menggunakan SVR pada jenis pelatihan  $\epsilon$ -SVR dan  $\nu$ -SVR, dan untuk mengetahui kombinasi fitur dengan hasil terbaik.

## 2 TINJAUAN PUSTAKA

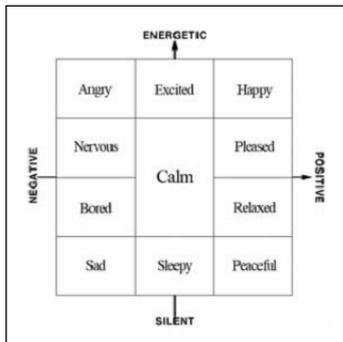
### 2.1. *Music Emotion Recognition*

Pada awalnya, banyak penelitian terkait dengan *mood* dan emosi dalam musik yang mempelajari tentang apa saja dan bagaimana pengaruh faktor *psychological* dan *physiological* yang berpengaruh pada pengenalan dan klasifikasi emosi. Pada tahun 1980-an beberapa model emosi diajukan dimana banyak diantaranya disusun berdasarkan pada pendekatan dimensional terhadap *emotion rating* [4].

Pendekatan dimensional mengidentifikasi jenis emosi berdasarkan pada posisinya terhadap beberapa dimensi emosi dengan sumbu tertentu dimana diharapkan dapat berkorespondensi dengan internal representasi emosi manusia [16]. Meskipun namanya berbeda, namun banyak peneliti yang mengajukan beberapa interpretasi emosi yang sebenarnya memiliki kemiripan [16].

Thayer menyarankan sebuah model emosi dalam bidang dua dimensi yang sederhana namun tangguh dalam mengorganisasi respon emosi yang berbeda

yaitu stress (*valence*) dan energi (*arousal*) [13]. Berdasarkan permodelan emosi dari Thayer dihasilkan permodelan emosi lain [4] yang menggunakan sebelas emosi berdasarkan teori Juslin. Permodelan emosi tersebut dapat dilihat pada Gambar 1.



**Gambar 1 Model emosi dua dimensi Thayer yang sudah dimodifikasi (Han, Rho, Dannenberg, & Hwang, 2009)**

Disebutkan oleh Han et al. bahwa banyak penelitian yang menginvestigasi pengaruh dari faktor musik seperti *loudness* dan *tonality* terhadap *perceived emotional expression*. Dalam menganalisa faktor tersebut digunakan metode yang bermacam-macam mulai, beberapa melibatkan pengukuran korelasi psikologis dan fisiologis antara keadaan dari keterangan faktor musikal dan pembangkitan emosi [4]. Juslin dan Sloboda menginvestigasi penggunaan dari *acoustic cues* dalam komunikasi dari emosi musik oleh pemain dan pendengar dan mengukur korelasi diantara ekspresi emosi (seperti kemarahan, kesedihan, kebahagiaan) dan *acoustic cues* (seperti *tempo*, *spectrum* dan artikulasi) [6].

Dalam pengenalan emosi, dapat dilakukan pembacaan emosi dengan memanfaatkan sistem koordinat. Sistem koordinat yang dapat diterapkan pada *emotion plane* diantaranya koordinat kartesius dan koordinat polar. Konversi koordinat kartesius menuju polar ataupun sebaliknya serta pengaplikasiannya terhadap

*emotion plane* ditunjukkan pada persamaan (1) dan persamaan (2) berikut:

$$E_c = (Ar, Vl), E_p = (Dt, An)$$

$$Dt = (Ar^2 + Vl^2)^{\frac{1}{2}}, An = \arctan\left(\frac{Ar}{Vl}\right) \quad (1)$$

$$Ar = Dt \cdot \cos(An), Vl = Dt \cdot \sin(An) \quad (2)$$

Keterangan:

$E_c$  = Titik pusat emosi pada sistem koordinat kartesius

$E_p$  = Titik pusat emosi pada sistem koordinat polar

$Vl$  = nilai *valence* (searah sumbu y pada kartesius)

$Ar$  = nilai *arousal* (searah sumbu x pada kartesius)

$Dt$  = jarak titik polar dari sumbu 0

$An$  = besar sudut yang dibentuk titik polar terhadap sumbu 0

## 2.2. Musical Features

### 2.2.1. Key Clarity

Fitur *key* merupakan bagian dari *harmony features* bersama dengan *musical mode*. Fitur ini didapatkan dengan mencari *key* atau nada yang dominan dari data audio. Metode yang digunakan dalam mencari *key* dari sebuah data musik adalah algoritma *key-finding* oleh Krumhansl.

Pencarian dilakukan dengan membandingkan *key profile matrix* dengan fitur *chroma*. Fitur *chroma* dinilai dapat merepresentasikan frekuensi pada skala musikal. Fitur *chroma* dan *key profile matrix* tersebut digunakan untuk mendapatkan nilai estimasi *tonality* dari data audio. Pencarian *tonality* dituliskan pada persamaan (3) sementara *key* terbaik dituliskan dengan persamaan (4).

$$\text{Tonality} = C \cdot \text{KeyProfileMatrix} \quad (3)$$

$$\text{Key} = \max_{\text{KeyIndex}} (\text{Tonality}(\text{Idx})) \quad (4)$$

Keterangan:

$Tonality$  = *tonality* dari sinyal  
 $C$  = nilai *chroma* dari sinyal

$KeyProfileMatrix$  = *key profile matrix* dari Krumhansl

$Key$  = key terbaik.

$Key\ clarity$  didapatkan dengan memilih *key* terbaik, yang merupakan *key* dengan nilai tertinggi pada *tonality*.

### 2.2.2. Musical Mode

Fitur *musical mode* merupakan bagian dari *harmony features*. Fitur *musical mode* sendiri diestimasi menggunakan nilai selisih dari *key* terbaik pada tangga nada mayor dan *key* terbaik pada tangga nada minor [16].

Untuk menghasilkan fitur *musical mode* dapat dilakukan pendekatan nilai *tonality* berdasarkan subbab sebelumnya. Setelah didapatkan nilai *tonality*, kemudian ditentukan nilai *key* terbaiknya pada masing-masing tangga nada untuk kemudian dapat dicari nilai selisihnya.

### 2.2.3. Zero Crossing Rate

*Zero crossing* adalah suatu keadaan di mana suatu fungsi menyentuh titik nol atau saat dimana suatu fungsi berpindah dari nilai positif ke negatif. *Zero crossing rate* dari data musik merupakan rasio terjadinya *zero crossing* dalam data musik. Metode ini telah banyak digunakan dalam pengenalan suara maupun informasi jenis musik.

Dalam konteks sinyal diskrit, *zero crossing* dapat terjadi jika sampel mempunyai tanda aljabar yang berbeda. Ini menunjukkan bahwa ada hubungan antara *zero crossing* dan konten frekuensi sinyal [15].

### 2.2.4. Pitch

*Pitch* disebut juga *perceived fundamental frequency*. Dalam menentukan nilai *pitch* terdapat banyak metode yang dapat digunakan, salah satunya yaitu

menggunakan metode autokorelasi. Autokorelasi merupakan formulasi sistematis yang digunakan dalam *signal processing* untuk menganalisa suatu fungsi domain waktu dari suatu sinyal maupun fungsi berbentuk deret. Nilai autokorelasi suatu sinyal akan menunjukkan bagaimana bentuk gelombang itu membentuk suatu korelasi dengan sinyal itu sendiri. Bentuk-bentuk pola yang sama atau mirip pada waktu tertentu menunjukkan perulangan bentuk atau perioditas pola sinyal suara. Autokorelasi ditunjukkan pada persamaan (5).

$$y_u(n) = \sum_{k=1}^M u(k)u(k+n) \quad (5)$$

Keterangan:

$y_u$  = hasil autokorelasi dari sinyal masukan  $u$

$n$  = lag parameter

$M$  = panjang sinyal masukan  $u$

### 2.2.5. Energy

*Energy features* berkaitan erat dengan persepsi arousal dari musik. Fitur *energy* yang dapat diekstrak salah satunya yaitu *Average Energy*. *Average energy* (AE) dari keseluruhan *wave sequence* diadopsi secara luas untuk mengukur *loudness* dari musik. Sementara regularitas dari *loudness* diukur melalui standar deviasi ( $\sigma$ ) dari AE. Perhitungan regularitas *loudness* ditunjukkan pada persamaan (6) sementara AE ditunjukkan pada persamaan (7).

$$\sigma(AE(x)) = \sqrt{\frac{1}{N} \sum_{t=0}^N (AE(x) - x(t))^2} \quad (6)$$

$$AE(x) = \frac{1}{N} \sum_{t=0}^N x(t)^2 \quad (7)$$

Keterangan :

$AE(x)$  = *Average Energy* atau nilai rata-rata dari sinyal  $x$

$N$  = panjang sinyal  $x$  pada sampel

$x(t)$  = nilai sinyal  $x$  pada waktu sampel  $t$

$t$  = waktu dalam sampel

$\sigma(AE(x))$  = standar deviasi dari sinyal  $x$ .

### 2.2.6. Tempo

*Average tempo* atau *global tempo* didefinisikan sebagai *tempo* atau jumlah *beat* tiap menit (BPM) dari sebuah musik. Fitur ini dapat mendiskripsikan ritme musik secara global.

Estimasi nilai *tempo* dapat dilakukan dengan berbagai macam pendekatan, salah satunya dengan menggunakan *onset strength*. *Onset strength* merupakan bentuk transformasi data musik yang berisikan informasi terkait *onset* pada data musik.

Untuk membentuk *onset strength* dari data musik dapat dilakukan dengan mengkomputasi *db mel spectrogram* dari data musik. Langkah dilanjutkan dengan membuang nilai negatif dari data *spectrogram* tersebut sehingga dihasilkan *onset* kasar. Kemudian dilakukan komputasi ulang terhadap *onset* kasar tersebut dengan menghitung nilai differensial orde pertamanya untuk membentuk *onset* baru dalam bentuk *halfwave*. Langkah terakhir yang dilakukan adalah membuang data *offsets* dengan *highpass filter* pada 0.5Hz.

Setelah dihasilkan *onset strength* dari data musik, langkah pencarian *global tempo* dilanjutkan dengan melakukan autokorelasi terhadap *onset strength* dan mengaplikasikan *preference window* terhadap data tersebut. Metode *windowing* yang digunakan adalah Gaussian pada *log-time axis*. Nilai *global tempo* dipilih dari *window* dengan nilai hasil autokorelasi terbesar.

### 2.2.7. Beat

Untuk dapat menghasilkan fitur *beat*, data musik perlu dikomputasikan *onset strength* dan *global tempo*-nya terlebih dahulu. Menggunakan nilai-nilai tersebut dilakukan proses *tracking* atau pencarian *beat* pada data. *Tracking* dilakukan dengan mengaplikasikan *dynamic programming*.

*Dynamic Programming* atau disebut juga sebagai *dynamic optimization* merupakan metode penyelesaian suatu permasalahan yang kompleks dengan memecahnya menjadi koleksi dari sub permasalahan yang lebih sederhana, kemudian menyelesaikan permasalahan-permasalahan tersebut sekali dan menyimpan semua solusinya.

Proses pencarian *beat* akan menghasilkan kumpulan *beat* yang kemudian dapat menghasilkan *tempo regularity* dengan mengkomputasi nilai standar deviasi dari interval *beat* tersebut.

## 2.3. Support Vector Regression

*Support Vector Regression* (SVR) merupakan pengembangan SVM untuk menangani kasus regresi.

SVR memiliki dua tipe strategi pelatihan yaitu  $\epsilon$ -SVR dan  $\nu$ -SVR. Pada  $\epsilon$ -SVR, hasil estimasi terbaik dari nilai koefisien dalam fungsi diperoleh untuk nilai  $\epsilon$ -*insensitive loss function* yang minimal dikarenakan nilai  $\epsilon$  (dibaca 'epsilon') yang rendah menunjukkan nilai aproksimasi yang tinggi. Namun,  $\epsilon$ -SVR ini memiliki keterbatasan dimana nilai  $\epsilon$  harus didefinisikan sebelum proses pelatihan sehingga akan muncul permasalahan ketika menghadapi data yang sulit diantisipasi selang nilai dari  $\epsilon$ -nya. Strategi lain yaitu  $\nu$ -SVR. Pada  $\nu$ -SVR ini, kekurangan dari  $\epsilon$ -SVR dapat diatasi dengan membatasi tugas dalam menemukan  $\epsilon$  untuk optimasi kuadratik dengan memperkenalkan parameter baru yaitu  $\nu$  (dibaca 'nu') yang berfungsi mengontrol jumlah *support vector* dan *training error*. Solusi akhir dari SVR *predictor* ditunjukkan pada persamaan (8) berikut:

$$f(x) = \sum_{i=1}^{\ell} (\alpha_i^* - \alpha_i) K(x_i, x) + b \quad (8)$$

Keterangan:

$f(x)$  = hyperplane

$\alpha_i^*$ ,  $\alpha_i$  = data points, langrange multipliers

$K$  = fungsi kernel

$b$  = bias

## 2.4. Mean Squared Error

*Mean Squared Error* (MSE) merupakan salah satu metode pengukuran *error* yang sering dipakai untuk menilai performansi suatu fungsi prediksi. Pada penelitian ini, nilai MSE dihitung berdasarkan target dari suatu data uji dimana nilai tersebut dibandingkan dengan hasil prediksi yang dihasilkan oleh model *Support Vector Regression*.

Rumus MSE sendiri dapat dituliskan dengan persamaan (9) sebagai berikut:

$$MSE = \frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2 \quad (9)$$

Keterangan:

MSE = *Mean Squared Error*

$m$  = Jumlah data

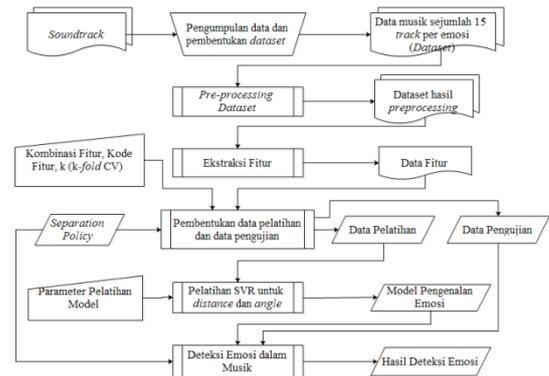
$\hat{y}_i$  = nilai output aktual data ke- $i$

$y_i$  = nilai prediksi untuk data ke- $i$

## 3 METODE PENELITIAN

Garis besar tahapan penelitian ini dapat dilihat pada Gambar 2. Langkah diawali dengan melakukan pengumpulan data dan pembentukan dataset. Langkah tersebut menghasilkan data musik yang akan digunakan sebagai dataset penelitian. Langkah dilanjutkan dengan melakukan *preprocessing* terhadap dataset yang terkumpul.

Dataset yang telah melalui tahapan *preprocessing* akan menjalani proses ekstraksi fitur yang akan menghasilkan data ciri atau fitur dari dataset. Berdasarkan data fitur tersebut akan dibentuk data pelatihan dan data pengujian.



Gambar 2 Tahapan Proses Pengenalan Emosi dalam Musik

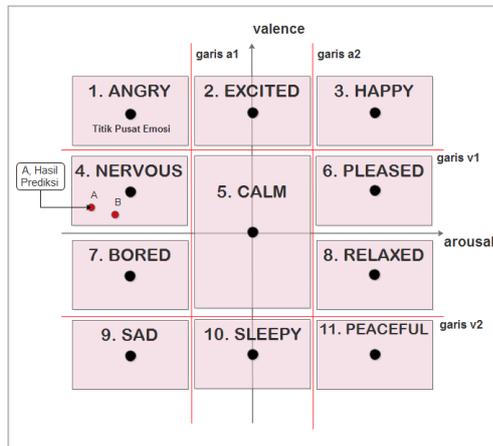
Dalam tahapan pembagian data tersebut, digunakan metode *k-fold cross validation*.

Data penelitian akan digunakan dalam proses pelatihan SVR yang bertujuan untuk membentuk model pengenalan emosi. Model yang dihasilkan berjumlah dua model yang nantinya akan mengenali titik koordinat polar dari data. Satu model digunakan untuk mengenali sudut (*angle*) dan model lainnya digunakan untuk mengenali jarak dari titik nol (*distance*). Sementara data pelatihan akan digunakan untuk menguji performa model yang dihasilkan dalam proses pelatihan untuk mengenali emosi data musik. Pada proses pengenalan emosi, emosi data yang dihasilkan berasal dari area emosi dari titik koordinat polar data emosi yang dihasilkan oleh model SVR tersebut.

Pada proses pelatihan maupun pengenalan diperlukan *groundtruth*. *Groundtruth* dibentuk dengan memanfaatkan *emotion plane* yang memetakan emosi. Emosi dipisahkan dengan fungsi *separation policy* berupa garis-garis yang membagi area *emotion plane* dan menghasilkan area untuk masing-masing emosi yang berjumlah sebelas emosi. *Groundtruth* pelatihan didapatkan dengan mencari titik pusat dari area emosi yang terbentuk. Titik pusat emosi tersebut dipresentasikan dengan koordinat polar (*angle, distance*). Hasil pengenalan dari model yang terbentuk dalam koordinat polar tersebut diaplikasikan kembali pada

*emotion plane* guna mendapatkan kategori emosinya.

Ilustrasi penggunaan *emotion plane* sebagai *groundtruth* dapat dilihat pada Gambar 3.



Gambar 3 Ilustrasi Penggunaan *Emotion Plane* sebagai *Groundtruth*

### 3.1. Pengumpulan Data dan Pembentukan *Dataset*

Jenis musik yang dianalisa dikhususkan pada musik *original soundtrack*. Data musik berupa musik pengiring yang digunakan pada film dengan mengutamakan jenis musik instrumental. Data didapatkan melalui beberapa situs pribadi di internet dan Youtube.

Dikarenakan masih minimnya kategorisasi emosi pada jenis data yang digunakan, pemilihan data musik yang digunakan sebagai dataset dilakukan dengan bantuan beberapa responden untuk memberikan pertimbangan terkait emosi yang terkandung dalam musik tersebut. Responden diminta untuk mendengarkan data musik instrumental dan diminta untuk memberi penilaian terkait kategori emosi yang dikandung data musik tersebut. Kegiatan tersebut bertujuan mendapatkan kategori emosi dari data musik.

Total emosi yang dikenali adalah 11 emosi, yaitu *angry*, *excited*, *happy*, *nervous*, *pleased*, *bored*, *calm*, *relaxed*, *sad*, *sleepy*, dan *peaceful*.

### 3.2. *Dataset Preprocessing*

Tiap-tiap data pada dataset perlu melewati beberapa tahapan data *preprocessing* untuk dapat diproses pada tahapan berikutnya. Tahapan data *preprocessing* ini bertujuan untuk menghasilkan *dataset* penelitian yang seragam dan seimbang.

Tahapan *preprocessing* diawali dengan melakukan pemotongan data audio. Panjang data audio yang digunakan adalah 25 detik. Setelah dilakukan pemotongan data kemudian dilakukan konversi format data, konversi *channel* dan *resampling*. Format data yang dituju adalah format *wave* (.wav) dengan *sampling rate* 22.050 Hz dan *channel* mono.

### 3.3. Ekstraksi Fitur

Ekstraksi fitur dilakukan terhadap *dataset* yang telah terbentuk dan melalui tahapan *preprocessing*. Tahapan ini dilakukan dengan tujuan mengambil ciri musikal dari data musik dan menghasilkan data fitur yang dapat diolah pada proses selanjutnya. Fitur-fitur yang diekstrak yaitu *energy*, *beat*, *tempo*, *key*, *musical mode*, *pitch*, dan *zero crossing rate (ZCR)*.

Proses ekstraksi fitur dilakukan dengan memanfaatkan fungsi-fungsi yang terdapat pada *library* Python Librosa, terutama untuk proses ekstraksi fitur *tempo*, *beat* dan *ZCR*.

### 3.4. Pembentukan Data Latih dan Data Uji

Proses pembentukan data pelatihan dan data pengujian bertujuan untuk mempersiapkan dan membentuk data fitur untuk dapat digunakan dalam proses *training* maupun *testing*.

Langkah diawali dengan memotong data fitur sesuai dan proses dilanjutkan dengan penambahan nilai target berupa *angle* dan *distance*. *Angle* dan *distance* merupakan

titik pusat emosi ditetapkan berdasarkan *groundtruth*.

Kemudian data dibagi menggunakan metode *k-fold cross validation* sehingga dihasilkan data latih dan data uji.

Terakhir dilakukan normalisasi. Normalisasi dilakukan pada data latih menghasilkan data ternormalisasi dan juga parameter normalisasi yang diperlukan untuk melakukan normalisasi pada data uji.

### 3.5. Pelatihan Model

Proses pelatihan model SVR dilakukan berdasarkan masukan data latih ternormalisasi dan parameter pelatihan model.

Jenis pelatihan SVR yang dilakukan yaitu  $\epsilon$ -SVR dan  $\nu$ -SVR. Pada tiap proses pelatihan, terdapat dua model yang dilatih yaitu model *angle* dan model *distance*.

### 3.6. Pengenalan Emosi dan Evaluasi

Proses pengenalan emosi dilakukan berdasarkan masukan data uji hasil pembagian data yang telah dinormalisasi. Data uji kemudian dikenali menggunakan model yang telah dilatih pada proses pelatihan SVR. Hasil prediksi berupa nilai *angle* dan *distance*.

Nilai *angle* dan *distance* digunakan dalam pencarian emosi berdasarkan pembagian emosi pada *separation policy*. Hasil proses tersebut kemudian dikenali sebagai emosi dari data uji. Hasil prediksi yang dihasilkan kemudian dianalisa akurasi berdasarkan persamaan . Hasil akurasi inilah yang menjadi tolak ukur kinerja dari model yang dilatih.

## 4 HASIL DAN PEMBAHASAN

### 4.1. Data Penelitian

Data penelitian yang digunakan adalah data audio *soundtrack* dari animasi televisi

atau film animasi yang didapatkan melalui beberapa situs di internet seperti Youtube. Data *soundtrack* yang dipilih berupa musik instrumental yang tidak diiringi dengan vokal. *Genre* dari *soundtrack* tidak dibatasi hanya pada satu *genre* musik tertentu.

Data musik tersebut dikumpulkan dan dipilih secara manual. Dipilih sejumlah 15 data musik untuk mewakili satu emosi sehingga total data yang dikumpulkan adalah sejumlah 165 data musik untuk mewakili keseluruhan 11 emosi yang digunakan.

Tiap data musik terpilih yang digunakan dipotong sepanjang 25 detik secara manual dan disimpan dalam file ekstensi (\*.wav). Data tersebut disimpan dalam bentuk *monochannel* dengan nilai *sampling rate* 22.050 Hz.

### 4.2. Skenario Pengujian

Dalam penelitian ini dilakukan beberapa skenario penelitian yang berfokus pada beberapa parameter yang terlibat dalam penelitian. Skenario penelitian dilakukan selain untuk mencari hasil terbaik dari pengenalan emosi, tetapi juga untuk menganalisa hasil pengenalan emosi dan pengaruh dari parameter-parameter yang terlibat didalamnya.

Skenario yang dilakukan diantaranya skenario pengujian 1 yang berfokus pada pencarian *separation policy* yang sesuai untuk data, skenario pengujian 2 yang berfokus pada pencarian kombinasi fitur terbaik dalam pengenalan emosi dan skenario pengujian 3 yang berfokus pada pencarian nilai terbaik untuk parameter pelatihan model *support vector regression*.

### 4.3. Hasil dan Analisa Skenario

#### 4.3.1. Hasil dan Analisa Skenario Pengujian 1

Berdasarkan skenario pengujian yang dilakukan dihasilkan *separation policy* yang ditunjukkan pada Tabel 1 dan Tabel 2.

**Tabel 1. Hasil *Separation Policy* terpilih  $\epsilon$ -SVR**

<i>Separation Policy</i>					Akurasi	Kode
v1(+)	v2(-)	a1(-)	a2(+)	V=0		
0,5	-0,5	-0,25	0,45	0	34,55%	$\epsilon$ -SP1
0,5	-0,5	-0,3	0,4	0	34,55%	$\epsilon$ -SP2

\*v artinya *valence* dan a artinya *arousal*, lebih jelas lihat pada Gambar 2.

**Tabel 2. Hasil *Separation Policy* terpilih v-SVR**

<i>Separation Policy</i>					Akurasi	Kode
v1(+)	v2(-)	a1(-)	a2(+)	V=0		
0,4	-0,5	-0,33	0,33	0	29,70%	v-SP1
0,45	-0,5	-0,33	0,33	0	29,09%	v-SP2

\*v artinya *valence* dan a artinya *arousal*, lebih jelas lihat pada Gambar 2.

Dimana berdasarkan skenario ini diketahui bahwa pencarian kombinasi *separation policy* ini memiliki peran penting dalam proses pengenalan emosi yang dilakukan. *Separation policy* tidak hanya menjadi dasar pembentukan *ground truth* sebagai target pelatihan, namun juga digunakan dalam klasifikasi emosi yang dilakukan.

#### 4.3.2. Hasil dan Analisa Skenario Pengujian 2

Berdasarkan skenario pengujian yang dilakukan dihasilkan kombinasi fitur dengan hasil akurasi terbaik yang ditunjukkan pada Tabel 3.

**Tabel 3. Akurasi Terbaik dari Kombinasi Fitur**

SP	Fitur Angle	Fitur Distance	Akurasi
$\epsilon$ -SP			
$\epsilon$ -SP1	<i>Musical Mode, Zero Crossing</i>	<i>Key Clarity, Zero Crossing</i>	34,55%

SP	Fitur Angle	Fitur Distance	Akurasi
	<i>Rate, Pitch, Energy, Tempo, Beat</i>	<i>Rate, Pitch, Energy, Tempo, Beat</i>	
$\epsilon$ -SP2	<i>Key Clarity, Zero Crossing Rate, Energy, Tempo</i>	<i>Key Clarity, Zero Crossing Rate, Energy, Beat</i>	35,76%
$\epsilon$ -SP2	<i>Key Clarity, Zero Crossing Rate, Energy, Tempo</i>	<i>Key Clarity, Zero Crossing Rate, Energy, Tempo</i>	35,76%
<b>v-SP</b>			
v-SP1	<i>Energy</i>	<i>Key Clarity</i>	36,97%
v-SP1	<i>Energy</i>	<i>Zero Crossing Rate</i>	36,97%
v-SP2	<i>Pitch, Energy</i>	<i>Zero Crossing Rate, Beat</i>	36,97%

Berdasarkan hasil yang didapatkan dapat dianalisa beberapa poin terkait hasil yang dicapai diantaranya sebagai berikut:

1. Pengaruh nilai MSE terhadap akurasi  
MSE merupakan parameter pengukuran *error* hasil prediksi dari model yang dihasilkan. Berdasarkan hal tersebut, parameter MSE ini dipilih sebagai parameter untuk memilih kombinasi fitur terbaik.

Dalam penggunaannya pada skenario yang telah dijalankan, kombinasi nilai MSE paling rendah tidak memastikan bahwa nilai akurasi yang dihasilkan merupakan akurasi paling baik. Namun, model dengan nilai MSE yang besar juga tidak menunjukkan performa yang baik.

Berdasarkan kasus tersebut, diketahui bahwa MSE merupakan parameter

berpengaruh pada identifikasi kelayakan model. Namun, MSE tidak menjadi parameter dalam mengukur akurasi model. Hal ini dikarenakan proses pengenalan emosi tidak hanya bergantung dari hasil prediksi yang dihasilkan satu model saja melainkan berdasarkan kombinasi nilai *angle* dan *distance* yang dihasilkan dan pemetaannya pada *emotion plane* berdasarkan *separation policy*.

## 2. Fitur

Hasil yang didapatkan menunjukkan bahwa pada pengenalan emosi menggunakan jenis pelatihan  $\epsilon$ -SVR maupun  $\nu$ -SVR, dengan melakukan pengkombinasian fitur dapat meningkatkan nilai akurasi. Meskipun pada jenis pelatihan  $\epsilon$ -SVR tidak dihasilkan peningkatan nilai akurasi yang signifikan, namun dapat diminimalisir jumlah fitur yang digunakan.

Pada jenis pelatihan  $\epsilon$ -SVR, kesemua fitur berpengaruh pada pengenalan emosi dengan kombinasi yang berbeda-beda. Sementara pada jenis pelatihan  $\nu$ -SVR, ditunjukkan fitur yang mempengaruhi pengenalan *angle* diantaranya *pitch* dan *energy*, sementara fitur yang mempengaruhi pengenalan *distance* yaitu *key clarity*, *zero crossing rate* dan *beat*.

*Angle* memiliki karakteristik nilai dalam notasi positif dan negatif, dimana nilai tersebut merepresentasikan besaran sudut dari titik emosi pada koordinat polar. Pada kuadran I dan II, nilai *angle* memiliki range nilai positif. Emosi pada kuadran I dan II diantaranya *angry*, *excited* dan *happy* akan cocok dengan musik yang memiliki nilai *pitch* yang tinggi/ kuat, musik yang berenergi dan *beat* yang cenderung cepat. Sementara pada kuadran III dan IV, nilai *angle* memiliki range nilai negatif. Emosi pada kuadran III dan IV diantaranya *sad*, *sleepy* dan *relaxed* yang sesuai dengan musik yang memiliki nilai *pitch* yang rendah, tidak terlalu

berenergi/ tidak terlalu menghentak-hentak dengan *beat* yang cenderung pelan.

Berbeda dengan *angle*, *distance* yang memiliki kisaran nilai dalam notasi positif dan merepresentasikan jarak titik terhadap titik pusat koordinat (0,0). Emosi yang memiliki jarak paling dekat dengan titik pusat koordinat yaitu *calm* dimana musik yang sesuai dengan karakteristik emosi tersebut adalah musik dengan *beat* dan *tempo* yang tidak tinggi maupun rendah dan cenderung stabil. Semakin jauh dari titik pusat, emosinya pun makin beragam diantaranya *angry*, *happy*, *sad* dan *peaceful*. Digambarkan bahwa musik yang memiliki nilai *tempo* yang rendah ataupun tinggi akan cenderung jauh dari titik pusat. Begitupun dengan fitur musik seperti *beat* dan *zero crossing rate*. Musik dengan *beat* tinggi yang cenderung sesuai dengan musik *happy* ataupun *angry* akan diidentifikasi lebih jauh dari titik pusat, begitupun musik dengan *beat* rendah seperti *sad* ataupun *peaceful*.

### 4.3.3. Hasil dan Analisa Skenario Pengujian 3

Berdasarkan skenario pengujian yang dilakukan dihasilkan kombinasi nilai parameter pelatihan model dengan hasil terbaik ditunjukkan pada Tabel 4 dan Tabel 5.

**Tabel 4. Hasil Pengujian Parameter pada  $\epsilon$ -SVR**

No	Cost		Gamma		Epsilon		Akurasi
	A	D	A	D	A	D	
1	2 <sup>6</sup>	2 <sup>6</sup>	2 <sup>-6</sup>	2 <sup>-7</sup>	10 <sup>-2</sup>	10 <sup>-3</sup>	35,55%
2	2 <sup>7</sup>	2 <sup>8</sup>	2 <sup>-7</sup>	2 <sup>-7</sup>	10 <sup>-2</sup>	10 <sup>-3</sup>	36,97%
3	2 <sup>7</sup>	2 <sup>8</sup>	2 <sup>-8</sup>	2 <sup>-7</sup>	10 <sup>-2</sup>	10 <sup>-2</sup>	36,97%

\*A artinya *angle*, D artinya *distance*

**Tabel 5. Hasil Pengujian Parameter pada  $\nu$ -SVR**

No	Cost		Gamma		Nu		Akurasi
	A	D	A	D	A	D	
1	2 <sup>7</sup>	2 <sup>7</sup>	2 <sup>-7</sup>	2 <sup>-6</sup>	2 <sup>-2</sup>	2 <sup>-2</sup>	36,97%
2	2 <sup>7</sup>	2 <sup>7</sup>	2 <sup>-7</sup>	2 <sup>-6</sup>	2 <sup>-2</sup>	2 <sup>-2</sup>	36,97%
3	2 <sup>7</sup>	2 <sup>7</sup>	2 <sup>-7</sup>	2 <sup>-6</sup>	2 <sup>-2</sup>	2 <sup>-2</sup>	37,58%

\*A artinya *angle*, D artinya *distance*

Berdasarkan hasil skenario pengujian yang dilakukan, perubahan nilai parameter pelatihan model memiliki pengaruh terhadap kinerja model yang diujikan. Pada pelatihan tipe  $\epsilon$ -SVR, terdapat perbedaan pada nilai cost dan nilai epsilon. Model *angle* dengan hasil baik memiliki nilai *cost* yang lebih kecil dan epsilon yang lebih besar. Berbeda dengan model *distance* dimana hasil baik didapatkan pada nilai *cost* yang lebih besar dan epsilon yang cenderung lebih kecil. Sementara nilai variabel  $\gamma$ nya cenderung bervariasi. Hal tersebut menunjukkan terdapat perbedaan kebutuhan terkait setting parameter model pada pelatihan  $\epsilon$ -SVR. Hal ini dapat dipengaruhi oleh nilai target pada *angle* dan *distance* yang memiliki perbedaan *range*. Berbeda halnya dengan hasil pengujian pada pelatihan tipe  $\nu$ -SVR. Pada pengujian ini, meskipun terdapat perbedaan namun nilai parameter yang muncul dari hasil pelatihan terbaik mirip, terutama pada parameter  $\nu$ .

Berdasarkan hal tersebut maka terlihat bahwa pada *angle* dan *distance* perlu dilakukan *setting* parameter dengan nilai yang berbeda.

#### 4.3.4. Analisa terhadap Hasil Pengenalan Emosi

Berdasarkan keseluruhan skenario yang dilakukan diambil hasil terbaik yang menghasilkan *confusion matrix* pada Tabel 6. Seperti yang terlihat pada *confusion matrix* Tabel 6 bahwa model hasil penelitian masih belum dapat mengenali keseluruhan emosi yang ada.

Hal ini terlihat jelas dikarenakan masih ada hasil emosi terprediksi yang kosong dan masih terkumpul pada beberapa emosi diantaranya dimana hasil paling tinggi terprediksi pada emosi *excited*, *pleased*, *relaxed* dan *sleepy*.

Tabel 6. *Confusion Matrix* Model Terbaik

		Predicted Class										$\Sigma$	
		1	2	3	4	5	6	7	8	9	10		11
Actual class	1	5	2	6	1	0	1	0	0	0	0	0	15
	2	0	6	2	0	0	5	0	1	0	0	1	15
	3	0	3	10	0	0	2	0	0	0	0	0	15
	4	0	1	0	0	0	6	0	5	0	1	2	15
	5	0	1	3	0	1	3	0	6	0	1	0	15
	6	0	1	1	0	0	9	0	4	0	0	0	15
	7	0	0	0	0	0	0	1	1	0	6	7	15
	8	0	0	1	0	0	2	0	10	0	0	2	15
	9	0	0	1	0	0	0	0	4	0	4	6	15
	10	0	0	0	0	0	0	0	2	0	10	3	15
	11	0	0	0	0	0	0	0	2	0	3	10	15
$\Sigma$		5	14	24	1	1	28	1	35	0	25	31	165

\*Class : <sup>1</sup>angry, <sup>2</sup>excited, <sup>3</sup>happy, <sup>4</sup>nervous, <sup>5</sup>calm, <sup>6</sup>pleased, <sup>7</sup>bored, <sup>8</sup>relaxed, <sup>9</sup>sad, <sup>10</sup>sleepy, dan <sup>11</sup>peaceful

Pemetaan emosi yang dihasilkan cenderung mengarah ke kuadran I dan kuadran IV.

Hasil akurasi yang rendah pada dua tipe pelatihan yang dilakukan dapat dipengaruhi oleh *dataset* yang digunakan dalam melakukan pengenalan emosi. Data *soundtrack* tidak memiliki *genre* khusus dalam penggunaannya sehingga meskipun data musik memiliki emosi yang sama namun *musical instrument* atau *genre* musiknya bisa saja berbeda. Hal tersebut menyebabkan data akan sulit dikenali dan diidentifikasi emosinya.

Selain itu, pemotongan data musik yang tidak tepat juga dapat mempengaruhi pengenalan emosinya. Meskipun pemotongan data telah dilakukan secara manual, namun masih dimungkinkan terjadinya kesalahan dalam memilih bagian yang mengimplikasikan emosi musik yang terkandung.

## 5 KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian mengenai Pengenalan Emosi dalam Musik dengan *Musical Features* menggunakan *Support Vector Regression* diantaranya:

1. Kombinasi fitur memiliki pengaruh terhadap pengenalan emosi dalam musik. Tiap-tiap kombinasi fitur menghasilkan model yang memiliki kinerja berbeda-beda dan mempengaruhi pengenalan emosi yang dilakukan baik pada model *angle* maupun model *distance*.
2. Parameter MSE berpengaruh terhadap mengukur kinerja model. Hasil penelitian menunjukkan bahwa nilai MSE dalam penelitian ini dapat digunakan dalam mengukur kinerja model meskipun tidak dapat digunakan sebagai identifikasi akurasi. Hal tersebut dikarenakan pengenalan emosi yang dilakukan tidak hanya berdasarkan hasil prediksi nilainya namun dipengaruhi letak hasil prediksi terhadap *separation policy* yang ditetapkan.
3. Parameter pelatihan model memiliki pengaruh dalam pembentukan model terbaik untuk mengenali emosi. Terkait dengan parameter pelatihan  $\epsilon$ -SVR dan  $\nu$ -SVR, masing-masing parameternya memiliki pengaruh yang berbeda terhadap hasil pelatihan. Parameter epsilon (atau  $\epsilon$ ) pada  $\epsilon$ -SVR merupakan parameter yang membatasi error dari hasil pelatihan perlu disesuaikan dengan kebutuhan model dimana hasil penelitian menunjukkan nilai parameter epsilon berbeda pada model *distance* dan *angle*. Untuk parameter  $\nu$ , model terbaik memiliki nilai  $\nu$  (atau  $\nu$ ) yang cenderung sama baik pada *angle* maupun *distance*, dimana nilai  $\nu$  yang

terlalu besar atau terlalu kecil akan cenderung menurunkan kinerja model.

4. Kombinasi terbaik yang diperoleh melalui penelitian ini dihasilkan menggunakan tipe pelatihan  $\nu$ -SVR dengan nilai akurasi yang tercapai sebesar 37,757%. Kombinasi fitur yang mencapai hasil tersebut yaitu fitur *Pitch* dan *Energy* pada model *Angle* dan *Zero Crossing Rate* dan *Beat* untuk model *Distance*. Kombinasi nilai parameter pelatihan model yang mencapai nilai tersebut ialah  $cost=2^7$ ,  $gamma=2^{-7}$  dan  $nu=2^{-2}$  untuk model *angle* dan  $cost=2^7$ ,  $gamma=2^{-8}$  dan  $nu=2^{-2}$  untuk model *distance*.

## DAFTAR PUSTAKA

- [1] Chang, C.-C., & Lin, C.-J. (2013, Maret 4). NTU CSIE. Dipetik Mei 17, 2019, dari LIBSVM -- A Library for Support Vector Machines: <https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>
- [2] Eerola, T., Lartillot, O., & Toivianen, P. (2009). Prediction Of Multidimensional Emotional Ratings In Music From Audio Using Multivariate Regression Models. *International Society for Music Information Retrieval 2009*, (hal. 621-626). Kobe.
- [3] Ellis, D. P., & Poliner, G. E. (2007). Identifying 'Cover Songs' with chroma feature and dynamic programming beat tracking. *IEEE Conf. on ICASSP*, vol. 4, 1429-1432.
- [4] Han, B.-j., Rho, S., Dannenberg, R. B., & Hwang, E. (2009). SMERS: Music Emotion Recognition using Support Vector Regression. *International Society for Music Information Retrieval*, (hal. 651-656).
- [5] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques Third Edition*. Waltham: Elsevier Inc.
- [6] Juslin, P. N., & Sloboda, J. A. (2001). *Music and Emotion: Theory and research*. Oxford Univ. Press.
- [7] Krumhansl, C. L. (1990). *Cognitive Foundation of Musical Pitch*. Oxford University Press.

- [8] MacDorman, K. F., Ough, S., & Chin-Chang, H. (2007). Automatic Emotion Prediction of Song Excerpts: Index Construction, Algorithm Design, and Empirical Comparison. *Journal of New Music Research* , 36.
- [9] Nalini, N. J., & Palanivel, S. (2015). Music Emotion Recognition: The Combined Evidence of MFCC and Residual Phase. *Egyptian Informatics Journal* , 1-10.
- [10] Refaailzadeh, P., Tang, L., & Liu, H. (2009). Cross Validation. Springer .
- [11] Rianto, J. (2011). Perangkat lunak Pengenalan Suara (Voice Recognition) Untuk Absensi Karyawan Dengan Menggunakan Metode Dynamic Time Warping (DTW). Bandung: Universitas Komputer Indonesia.
- [12] Smola, A. J., & Scholkopf, B. (2004). A Tutorial on Support Vector Regression. *Statistics and Computing* , 199-222.
- [13] Thayer, R. E. (1989). *The Biopsychology of Mood and Arousal*. Newyork: Oxford University Press.
- [14] Tian, M., Fazekas, G., Black, D. A., & Sandler, M. (2015). On the Use of Tempogram to Describe Audio Content and Its Application to Music Structural Segmentation. 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (hal. 419-423). London: IEEE.
- [15] Wikessa, P. S., Hidayat, B., & Atmaja, R. D. (2017). Perancangan Aplikasi Deteksi Kemacetan Berdasarkan Audio Processing Menggunakan Metode Zero Crossing Rate dan Average Energi Berbasis Android. *e-Proceeding of Engineering:Vol.4* , 412-419.
- [16] Yang, Y.-S., & Chen, H. H. (2011). *Music Emotion Recognition*. USA: CRC Press.
- [17] Zhang, Y., Wang, S., Philips, P., & Ji, G. (2014). Binary PSO with Mutation Operator for Feature Selection using Decision Tree Applied to Spam Detection. *Knowledge-Based Systems*.