

# WKLASIFIKASI EMOSI PADA LIRIK LAGU MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

## EMOTION CLASSIFICATION IN A SONG LYRICS USING SUPPORT VECTOR MACHINE METHOD

<sup>1</sup>Irfan Dhiya Abirawa, <sup>2</sup>Andrew Brian Osmond, <sup>3</sup>Casi Setianingsih

<sup>123</sup>Program Studi S1 Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom

<sup>1</sup>irfandabirawa@gmail.com, <sup>2</sup>abosmond@telkomuniversity.ac.id, <sup>3</sup>setiacasie@telkomuniversity.ac.id

### Abstrak

Lagu merupakan sebuah salah satu elemen yang paling berpengaruh dalam menentukan ekspresi dan emosi. Dengan elemen yang bersifat audio dan dapat di representasikan maknanya yang lebih kuat dalam lirik lagu.

Klasifikasi lagu dapat dilakukan dengan menggunakan lirik lagu sebagai media yang digunakan untuk mengklasifikasi ekspresi dan emosi seseorang.

Dengan menggunakan model *Emotion Ontology Thayer Model*, klasifikasi *text processing* digunakan dengan menerapkan *future part-of-speech* dalam proses *pre processing* dengan *Support Vector Machine* sebagai algoritma untuk mengklasifikasikan ekspresi dan emosi sebuah lagu. Diharapkan dengan menggunakan *Support Vector Machine* dapat meningkatkan akurasi hasil klasifikasi.

**Kata kunci :** Lirik Lagu, *Text Processing*, *part-of-speech*, *Support Vector Machine*.

### Abstract

Song is one of the most influential elements in determining expression and emotion. With an elements that are audio and can be represented more powerful meaning in the lyrics of the song.

Classification of songs can be done by using song lyrics as a medium used to classify expression and emotions a someone.

Using the *Emotion Ontology Thayer Model*, *text-processing classification* is used by applying a *part-of-speech future* in the *pre-processing process* with the *Support Vector Machine* as an algorithm to classify the expression and emotion of a song. It's hoped that using *Support Vector Machine* can improve the accuracy of classification results.

**Keywords :** *Song Lyrics*, *Text Processing*, *part-of-speech*, *Support Vector Machine*.

## 1. Pendahuluan

### 1.1 Latar Belakang

Musik merupakan sarana yang ampuh dan memiliki banyak kebaikan bagi tubuh dan jiwa manusia, musik dapat menenangkan bahkan membangkitkan semangat seseorang yang mendengarkan musik. Para ahli psikologi menyatakan bahwa terdapat bagian pada otak manusia yang merasakan musik dekat dengan bagian otak yang berhubungan dengan ekspresi emosional manusia. Maka dari itu terdapat hubungan langsung antara musik dan emosi [1].

Terdapat dua kelompok peneliti yang mempelajari hubungan antara musik dengan emosi, yaitu ; 1) Psikolog Musik, mereka yang mempelajari hubungan antara isyarat akustik (*beat*, tempo, tingkat suara, dll.) dengan berbagai ekspresi emosi (marah, sedih, senang, tenang, dll.) dan sebagian besar mereka yang mengembangkan model emosional. 2) Peneliti bidang komputer, yaitu mereka yang mengembangkan algoritma untuk mendeteksi emosi terhadap musik secara otomatis [1] Mereka mencoba menggunakan emosi selain untuk mendapatkan *metadata* konvensional seperti jenis musik, musik mood, yang juga sebagai penelitian dalam bidang *Music Information Retrieval*.

Sekarang banyak sekali musik digital yang beredar di kalangan masyarakat. Berdasarkan catatan Billboard, penjualan musik digital hamper sekitar 50,3 persen dari penjualan musik di tahun 2011 dan 2012. Sementara itu, penjualan musik berbentuk fisik menurun hingga 2 tahun ke depan. Penjualan single

dan album digital meningkat sehingga memberikan kontribusi sebesar 34 persen terhadap total pendapatan industri musik [2]

Berdasarkan informasi diatas bahwa pemanfaatan musik digital meningkat. Lagu-lagu yang beredar berdasarkan ekspresi dan emosi yang dirasakan pengguna adalah suatu ide yang dapat dikembangkan untuk mengetahui penggunaan musik digital terhadap pengguna sehingga pengguna dapat memilih lagu tidak berdasarkan gender musik yang diminati melainkan berdasarkan suasana hati mereka.

Sebelumnya sudah ada sistem yang dapat menentukan judul lagu yang ingin di dengarkan pengguna dan mengetahui *gender* dari penulis lagu tersebut tetapi belum ada sistem yang dapat mengklasifikasi emosi apa yang terdapat pada lagu tersebut. Oleh karena itu dibuatlah sistem pengklasifikasian emosi berdasarkan lirik lagu.

Untuk mengklasifikasikan emosi yang terdapat pada lirik lagu tersebut menggunakan metode *Support Vector Machine* dan pembobotan menggunakan TF-IDF. Dengan metode *Support Vector Machine* tersebut dapat menentukan klasifikasi emosi apakah yang sesuai pada lirik lagu tersebut dan dengan metode ini juga diharapkan bisa mendapat hasil yang tingkat keakuratannya tinggi.

## 1.2 Tujuan

Adapun tujuan yang akan dicapai pada Tugas Akhir ini :

Mengklasifikasikan lirik lagu yang terdapat pada musik digital terhadap emosi pengguna sehingga dapat menentukan lagu apa yang sesuai dengan suasana hati mereka. Berdasarkan survei yang kami lakukan bahwa banyak sekali yang membutuhkan pengklasifikasian emosi berdasarkan lagu yang didengar oleh penikmat musik digital, sehingga dibuat lah sistem pengklasifikasian emosi berdasarkan lirik lagu.

## 1.3 Identifikasi Masalah

Beberapa identifikasi masalah yang akan dibahas dalam Tugas Akhir ini adalah sebagai berikut :

Rumusan masalah yang terdapat di dalam tugas akhir ini yaitu sulitnya mengklasifikasi emosi sebuah lagu yang di dengarkan oleh pengguna . Penelitian-penelitian sebelumnya yang melakukan klasifikasi emosi berdasarkan lirik lagu hanya untuk lagu berbahasa Inggris. Pada penelitian ini adalah penyempurnaan pada penelitian-penelitian sebelumnya dengan lirik lagu berbahasa Indonesia dan mempunyai tahapan *preprocessing* dengan metode *Support Vector Machine* (SVM).

# 1. Dasar Teori

Bagian ini berisi tentang dasar teori yang digunakan untuk merancang sistem Klasifikasi emosi berdasarkan lirik lagu.

## 1.1 Teks

Teks menurut wacana berarti lisan yang difiksasikan dalam bentuk tulisan. Teks juga dapat diartikan sebagai seperangkat tanda yang ditransmisikan dari seorang pengirim kepada seorang penerima melalui medium tertentu. Teks juga bisa diartikan satuan lingual yang dimediasi secara tulis atau lisan.

Struktur teks adalah tata organisasi teks, yaitu cara teks disusun. Jenis-jenis teks ada macam-macam seperti surat, editorial, iklan, negosiasi, naratif, dan lain-lain. Lalu teks memiliki struktur teks yang berbeda seperti, jenis verba, konjungsi, kelompok kata, dan lain-lain.

## 1.2 Text Mining

*Text mining* merupakan penerapan konsep dan teknik data *mining* untuk mencari pola dalam teks, yaitu proses penganalisisan teks guna mencari informasi yang bermanfaat untuk tujuan tertentu. Berdasarkan ketidakaturan struktur data teks, maka proses teks *mining* memerlukan beberapa tahap agar teks dapat diubah menjadi lebih terstruktur. *Preprocessing* adalah salah satu metode yang digunakan untuk memisahkan kata dari kalimat dan digunakan sebagai *keyword*.

### 1.3 Preprocessing

*Preprocessing* adalah proses pengubahan bentuk data yang belum terstruktur menjadi data yang terstruktur. *Preprocessing* diperlukan untuk memilih kata dalam kalimat yang akan digunakan sebagai *keyword* atau indeks. Indeks ini akan digunakan sebagai kata yang mewakili suatu dokumen untuk membuat pemodelan teks *mining*. Adapun empat tahapan *preprocessing*, yaitu : *Case Folding, Filtering, Stemming, Tokenizing,* dan *Post Tagging*.

### 1.4 Proses Sistem Temu Kembali Informasi

Proses temu kembali teks yang lebih dikenal dengan nama *text information retrieval*, merupakan sebuah teknik pencarian dengan menggunakan algoritma tertentu untuk mendapatkan hasil pencarian yang relevan berdasarkan kumpulan (*corpus*) informasi yang besar. Sebagian besar penggunaan sistem temu kembali adalah pada teks. Pengguna memasukkan kata kunci berupa teks, dan kemudian sistem mengolahnya hingga mendapatkan informasi semantik yang diinginkan oleh pengguna. [7]

### 1.5 TF-IDF

Pembobotan adalah proses penghitungan bobot tiap *term*/kata yang dicari pada setiap dokumen sehingga dapat diketahui ketersediaan dan kemiripan suatu kata di dalam dokumen. Pada tugas akhir ini pembobotan yang digunakan menggunakan Pembobotan TF-IDF. *Term frequency (tf)* adalah frekuensi dari kemunculan sebuah *term* dalam dokumen yang bersangkutan. *Idf* merupakan sebuah perhitungan dari bagaimana *term* didistribusikan secara luas pada koleksi dokumen yang bersangkutan.

*Inverse document frequency* menunjukkan hubungan ketersediaan sebuah *term* dalam seluruh dokumen. Semakin sedikit jumlah dokumen yang mengandung *term* yang dimaksud, maka nilai *idf* semakin besar. Nilai *idf* sebuah *term* dirumuskan dalam persamaan berikut :

$$W_{dt} = t_{f_{dt}} * IDF_t$$

Keterangan :

d : dokumen ke-d  
 t : kata ke-t dari kata kunci  
 W : bobot dokumen ke-d terhadap kata ke-t  
 tf : banyaknya kata yang dicari pada sebuah dokumen  
 IDF :  $\log_2 (D/df)$   
 D : total dokumen  
 df : banyak dokumen yang mengandung kata yang dicari

### 1.6 Support Vector Machine

*Support Vector Machine (SVM)* dikembangkan oleh Boser, Guyon, dan Vapnik, pertama kali diperkenalkan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. Konsep dasar metode SVM sebenarnya merupakan gabungan atau kombinasi dari teori-teori komputasi yang telah ada pada tahun sebelumnya, seperti *marginhyperplane* [4] kernel diperkenalkan oleh Aronszajn tahun 1950, *Lagrange Multiplier* yang ditemukan oleh Joseph Louis Lagrange pada tahun 1766, dan demikian juga dengan konsep-konsep pendukung lain. [6].

### 1.7 Support Vector Machine dengan RBF Kernel

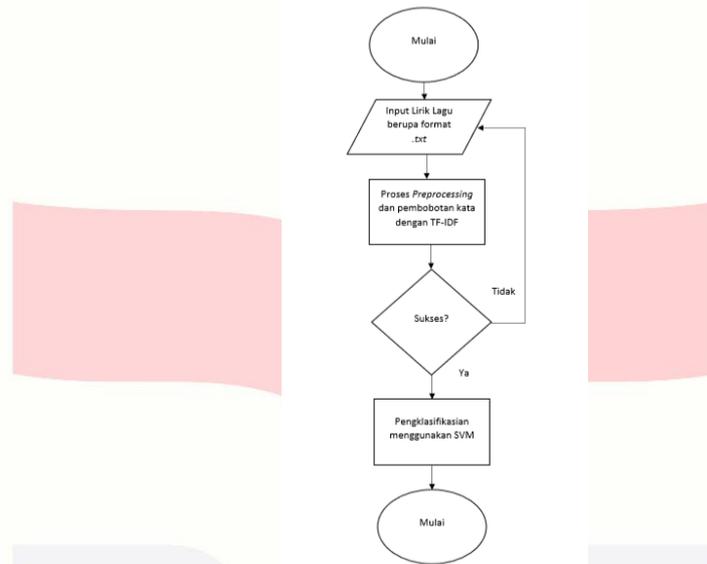
*Radial Basis Function* merupakan salah satu dari 3 kernel pada SVM. Susunan terdiri dari sebuah jarring RBF secara umum yang terdiri dari tiga lapisan. Lapisan Layer masukan terbentuk dari titik sumber. Lapisan kedua adalah lapisan hidden pada multilayer dan lapisan ketiga adalah hasil atau *output*.

Lapisan keluaran memberikan tanggapan dair lapisan masukan. Transformasi dari lapisan masukan ke lapisan hidden adalah nonlinear. Sedangkan transformasi dari lapisan hidden ke lapisan keluar adalah linear.

## 2. Perancangan

### 2.1 Perancangan Sistem Secara Umum

Pada perancangan sistem ini menjelaskan secara umum mengenai tahapan sistem yang akan diteliti lebih lanjut. Berikut merupakan perancangan sistem secara umum yang ditujukan pada gambar 2.1.



**Gambar 2.1 Diagram Alir Sistem pada Program**

Pada gambar 2.1 menjelaskan proses program dimulai dari menginputkan lirik lagu lalu dilakukan *preprocessing* dan pembobotan kata dengan TF-IDF. Lalu hasilnya akan di uji dan diklasifikasikan dengan algoritma *Support Vector Machine*. Hasil akan berupa label emosi dari lirik lagu tersebut.

## 3. Pengujian

### 3.1. Pengujian Black Box

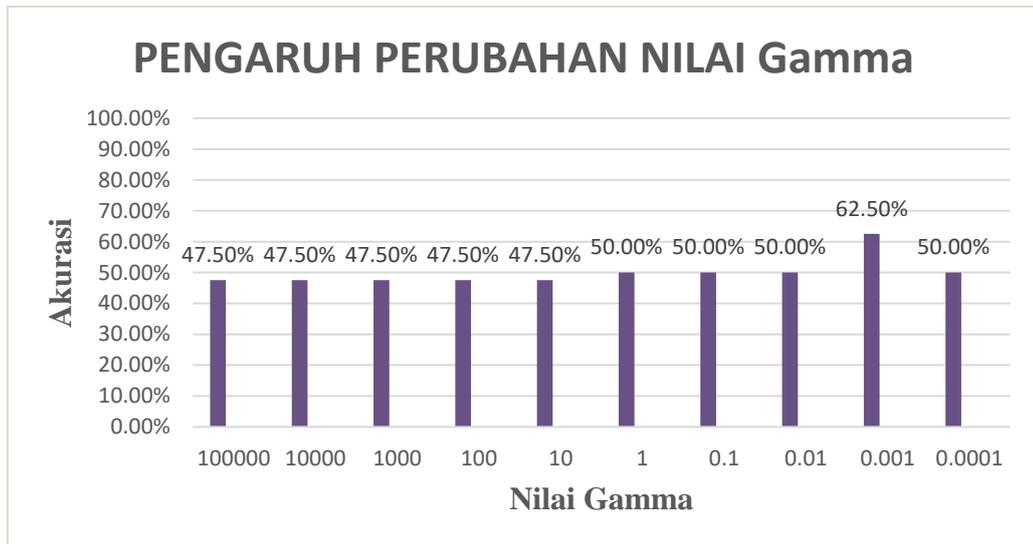
Sebelum dilakukan pemrosesan dilakukanlah pengujian aplikasi berguna memperlancar pengujian data testing. Adapun yang diuji adalah pengujian Black Box dimana aplikasi diuji dari awal sampai akhir.

**Tabel 3.1 Pengujian Black Box**

Hasil Pengujian				
No	Data Masuk	Estimasi	Realita	Kesimpulan
1	Menginputkan Lirik Lagu	Memasukan Data Lirik Lagu	Menuliskan Nama Data Lirik Lagu	Kurang Sempurna
2	Melakukan Preprocessing	Melakukan Preprocessing Lirik Lagu	Melakukan Proses Preprocessing	Sukses
3	Melakukan TF-IDF	Melakukan TF-Idf	Melakukan TF-Idf	Sukses
4	Melakukan SVM	Melakukan SVM	Mengklasifikasikan Label Emosi	Kurang Sempurna
5	Kosongkan Hasil	Menghapuskan Hasil	Menghapuskan Hasil	Sukses

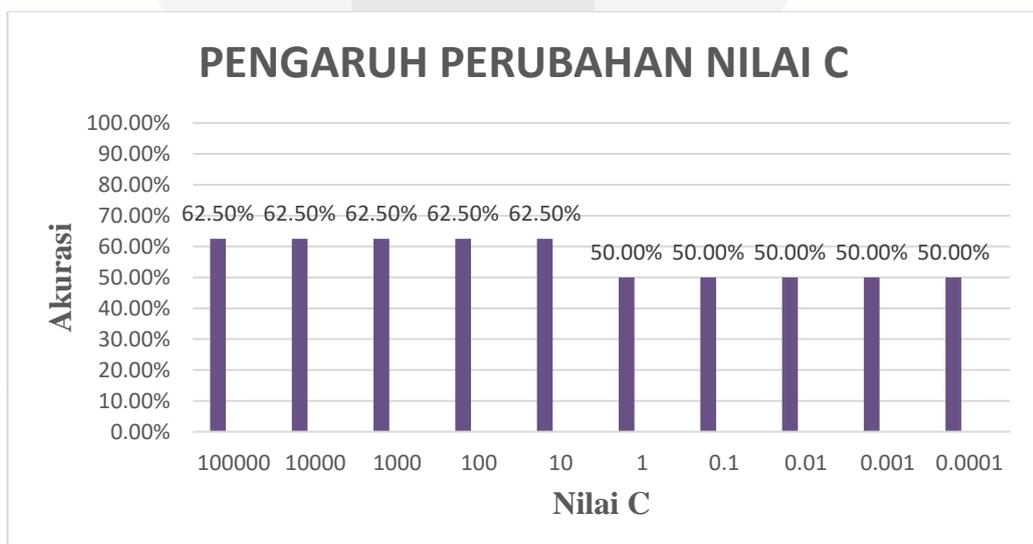
### 3.2. Pengujian Skenario

Hasil pengujian dari pengaruh nilai Gamma dari skala 1000, 100, 10, 1, 0.1, 0.01, 0.001, 0.0001, 0.00001, 0.000001 dengan nilai C tetap atau *Default* dan dengan hasil terbaik pada nilai 10 maka dapat kita lihat hasil yang memiliki nilai akurasi yang terbaik berada di nilai gamma = 0.0001 dengan nilai akurasi 62.50% setelah melewati nilai lain mengalami penurunan akurasi. Untuk mengetahui seberapa perbedaan antara perubahan nilai gamma, maka hasil dibuat dalam bentuk grafik seperti pada Gambar 3.1 dibawah ini :



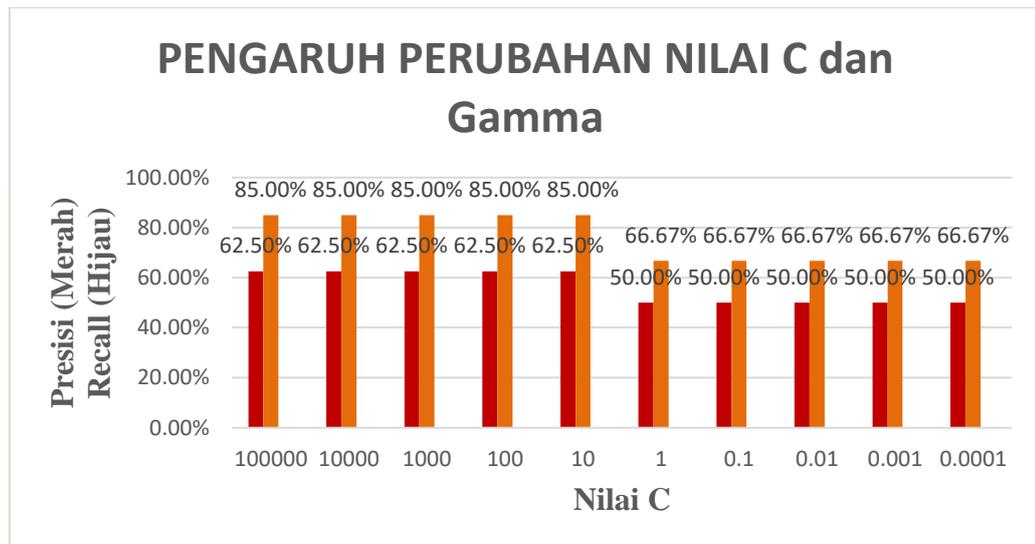
Gambar 3.1 Hasil Pengujian Nilai Gamma

Dari hasil pengujian pengaruh nilai C dari skala 100000, 10000, 1000, 100, 10, 1,  $10^{-1}$ ,  $10^{-2}$ ,  $10^{-3}$ ,  $10^{-4}$ , dengan nilai Gamma tetap atau *Default* dan dengan hasil terbaik pada nilai 0.001 maka dapat kita lihat hasil yang memiliki nilai akurasi yang terbaik berada di nilai C= 10 dengan nilai akurasi 62.50% setelah melewati nilai 10 keatas maka hasil akurasi akan mengalami kestabilan. Untuk mengetahui seberapa perbedaan antara perubahan nilai C, maka hasil dibuat dalam bentuk grafik seperti pada Gambar 3.2 dibawah ini :



Gambar 3.2 Hasil Pengujian Nilai C

Lalu untuk Presisi dan Recall terbaik juga sama yaitu nilai  $c$  dan  $\gamma$  diatas 10. Adapun grafik Presisi dan Recall bisa dilihat di Gambar 3.3 dibawah ini :



**Gambar 3.3** Presisi Dan Recall Pada Sistem

## 1. Kesimpulan

Dari skenario dan pengujian yang telah dilakukan dan dibahas sebelumnya dapat disimpulkan bahwa:

1. Pengklasifikasian emosi berdasar lirik lagu berbahasa Indonesia dapat dilakukan dengan memperbanyak data latih yang sudah dilabelkan emosi dengan kamus atau korpus Bahasa.
2. Hasil akurasi tertinggi dan terbaik didapatkan dari nilai  $c$  sebesar 10 atau selebihnya dan nilai  $\gamma$  sebesar 0.001 dengan akurasi sebesar 62.50%
3. Sistem yang dirancang berhasil mengklasifikasikan emosi berdasarkan lirik lagu berbahasa Indonesia secara otomatis.

## Daftar Pustaka:

- [1] Niaga Swadaya, "Cerdas dengan Musik" , 2014
- [2] IFPI, "IFPI Digital Music Report 2013 : Engine of a digital world", 2013.
- [3] Shigeo Abe, "Support Vector Machines for Pattern Classification" , *Springer Science & Business Media, Second Edition*, 2010
- [4] Richard O. Duda, Peter E. Hart, "Pattern Classification and Scene Analysis" , *Wiley*, 1973.
- [5] Sayed Fachrurrazi, "Penggunaan Metode Support Vector Machine (SVM) untuk Mengklasifikasi dan Memprediksi Angkutan Udara jenis Penerbangan Domestik dan Penerbangan Internasional di Banda Aceh" , *FMIPA USU Program Studi S2 Teknik Informatika, Thesis*, 2011.
- [6] Bernhard Schölkopf, Alexander J. Smola, "Learning with Kernels : Support Vector Machines, Regularization, Optimization, and Beyond", *MIT Press*, 2002

- [7] Yuita Arum Sari, Achmad Ridok, Marji, "PENENTUAN LIRIK LAGU BERDASARKAN EMOSI MENGGUNAKAN SISTEM TEMU KEMBALI INFORMASI DENGAN METODE LATENT SEMANTIC INDEXING", *Jurnal Seminar Nasional Teknologi Informasi dan Komputasi (SENASTIK)*, 2012.
- [8] Amelia Rahayu Dwi, "Klasifikasi Emosi Pada Lirik Lagu Dengan Metode Naïve Bayes Classifier (NBC) dan Logika Fuzzy", *Jurnal Tingkat Sarjana Jurusan Teknik Informatika Universitas Brawijaya*, 2013.
- [9] Mohammed Yasser Burhan, Marji, "PENGELOMPOKAN EMOSI BERDASARKAN LIRIK LAGU MENGGUNAKAN ALGORITMA ITERATIVE DICHOTOMIZER TREE", *Jurnal Tingkat Sarjana Jurusan Ilmu Komputer Universitas Brawijaya*, 2013.
- [10] Ali, Utsman, "Pengertian Emosi dan Bentuk Emosi".[Online]. <http://www.pengertianpakar.com/2015/03/pengertian-emosi-dan-bentuk-emosi.html> [Diakses 19 September 2017].
- [11] Sanjaya, Ade. "Pengertian Lagu Musik Definisi Pengaturan Menurut Para Ahli dan Hukum Undang Undang". [Online]. <http://www.landasanteori.com/2015/09/pengertian-lagu-musik-definisi.html> [Diakses 4 Oktober 2017].
- [12] Rocky Yefnes Dillak, "Klasifikasi Jenis Musik Berdasarkan File Audio menggunakan Jaringan Syaraf", *ISSN International Journal*, 2012.
- [13] F. Wulandini and A. S. Nugroho, "Text Clasification Using Support Vector Machine for Webmining Based Spatio Temporal Anlysis of the Spread of Tropical Diseases," International Conference on Rural Information and Communication Technology, pp. 189-192, 2009.