

KLASIFIKASI EMOSI PADA LIRIK LAGU MENGGUNAKAN METODE K-NEAREST NEIGHBOR

EMOTION CLASSIFICATION IN SONG LYRICS USING K-NEAREST NEIGHBOR METHOD

¹Afif Hijra Ferdinan, ²Andrew Briand Osmond S.T, M.T., ³Casi Setianingsih S.T., M.T.

^{1,2,3}Program Studi S1 Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom

¹afifhijraf@gmail.com, ²abosmond@telkomuniversity.ac.id, ³casie.sn@gmail.com

Abstrak

Setiap orang tentu pernah mengalami berbagai emosi. Emosi adalah perasaan seseorang terhadap suatu keadaan yang mempengaruhi tingkah laku. Emosi bersifat subyektif dan temporer yang muncul atau dipicu oleh stimulus berupa pengalaman dari orang-orang sekitar dan lingkungannya. Salah satu contoh lingkungan yang dapat merangsang emosi seseorang adalah dari lagu yang didengarkan.

Lagu adalah suatu kesatuan musik yang terdiri atas susunan nada-nada yang teratur dan terdapat lirik didalamnya. Nada-nada yang teratur disebut dengan melodi, dan lirik adalah syair atau kata-kata yang disuarakan mengiringi melodi. Lirik lagu adalah bagian yang berperan membangun emosi. Pemilihan kata-kata yang tepat diperlukan untuk menciptakan emosi yang sesuai.

Dalam Tugas Akhir ini, dibuktikan penulis dapat membuat sebuah sistem pengklasifikasian emosi berdasarkan lirik lagu. Dengan menggunakan metode *data mining* dengan algoritma *k-Nearest Neighbor*. Setelah dibuatnya buku dan sistem ini, maka sistem ini dapat digunakan sebagai alat pengklasifikasian emosi berdasarkan lirik lagu berbahasa Indonesia dengan hasil rata-rata akurasi sebesar 75%.

Kata kunci : Emosi, *Data Mining*, *K-Nearest Neighbor*

Abstract

Every human being is naturally experiences various emotions. Emotion is a person's feelings into something that affect person's behavior itself. Emotion is subjective and temporary that occurred or triggered by a stimulus such as experience from other people around him and the environment. One of the environment that can trigger someone's emotion is from the song being listened to.

Songs are music that consists of a regular arrangement of tones and there are lyrics inside. The tones are usually called melodies, and lyrics called as poem or words that are voiced together along melody. Song lyrics are parts that can build emotions. Choosing the right words for lyrics are very important because it will create the right emotion. Then, choosing the right songs to listen is also important.

In this final project, writer made a system that can classifying emotion from a song lyrics. Using one of the method in data mining which is k-Nearest Neighbor Algorithm. After the creation of this book, this system can be used as a classification of emotion based on Indonesian song lyrics within the average of accuracy is 75%

Keywords : *Emotion, Data Mining, K-Nearest Neighbor*

1. Pendahuluan

1.1 Latar Belakang

Setiap orang tentu pernah mengalami berbagai emosi. Emosi adalah perasaan seseorang terhadap suatu keadaan atau lingkungan sekitar yang mempengaruhi tingkah laku. Dalam kehidupan sehari-hari, penyampaian emosi dapat disampaikan secara non-verbal menggunakan ekspresi wajah maupun verbal berupa tingkah laku. Emosi bersifat subyektif dan temporer yang muncul atau dipicu oleh stimulus seperti perlakuan dari orang sekitar atau lingkungannya. Salah satu contoh lingkungan luar yang dapat merangsang emosi seseorang adalah dari lagu yang didengarkan [2]. Lagu sudah menjadi bagian dalam kehidupan manusia yang setiap hari mendengarkannya.

Lagu adalah suatu kesatuan antara musik yang terdiri atas susunan nada-nada teratur dan lirik. Susunan nada tersebut akan membentuk suatu irama dan lirik adalah syair atau kata-kata yang disuarakan mengiringi irama tersebut. Lirik lagu adalah salah satu bagian dalam lagu yang berperan membangun emosi. Pemilihan

kata-kata yang tepat diperlukan untuk menciptakan emosi yang sesuai [3]. Emosi yang terdapat pada lirik lagu akan menjadi stimulus bagi pendengarnya. Lirik lagu juga menunjukkan suasana hati seseorang. Dalam mengekspresikan emosinya, pencipta lagu menyusun kata-kata untuk dijadikan sebuah lirik lagu.

Seiring berkembangnya zaman, jumlah lagu yang tersedia di dunia semakin banyak. Salah satu stimulus luar untuk emosi pun semakin banyak, yaitu lagu itu sendiri. Berdasarkan *International Federation of the Phonographic Industry* (IFPI) perkembangan jumlah lagu pada tahun 2017 meningkat hingga 8,1%. Dengan 54% diantaranya berupa lagu digital. Penghasilan yang didapatkan dari lagu digital pun meningkat hingga 19,1% sedangkan penghasilan yang didapatkan dari musik berbentuk fisik menurun hingga 5,4%.

Dengan bertambahnya jumlah lagu digital maka diperlukannya suatu sistem yang dapat mengklasifikasikan lagu-lagu tersebut. Pengklasifikasian lagu berdasarkan emosi juga belum banyak digunakan. Padahal emosi cenderung berperan besar dalam komunikasi antar manusia di kehidupan sehari-hari. Terdapat beberapa metode *machine learning* yang dapat digunakan untuk klasifikasi. Pada tugas akhir ini penulis menggunakan metode *k-Nearest Neighbor* (k-NN). Pada penelitian k-NN sebelumnya yang digunakan untuk klasifikasi emosi untuk teks berbahasa Indonesia mencapai akurasi dengan tingkat 60% [5]. Dengan nilai keakuratan yang sudah cukup, maka bisa dibilang metode ini cocok digunakan untuk mengklasifikasikan lirik lagu.

1.2 Tujuan

Berdasarkan pada rumusan masalah diatas, maka tujuan yang dibahas pada tugas akhir ini adalah merancang sistem yang dapat mengklasifikasikan emosi berdasarkan lirik lagu secara otomatis agar pengguna dapat memilih lagu yang sesuai dengan suasana hatinya.

1.3 Identifikasi Masalah

Adapun identifikasi masalah berdasarkan latar belakang di atas adalah bagaimana mengklasifikasikan emosi pada lirik lagu secara akurat. Bagaimana cara merancang sistem yang dapat mengklasifikasikan emosi berdasarkan lirik lagu secara otomatis. Dengan harapan pengguna dapat memilih lagu yang sesuai dengan suasana hatinya berdasarkan hasil klasifikasi yang dihasilkan.

2. Dasar Teori

Bagian ini berisi tentang dasar teori yang digunakan untuk merancang sistem klasifikasi dengan mengimplementasikan metode k-Nearest Neighbor untuk klasifikasi emosi pada lirik lagu. Adapun teori-teori yang digunakan adalah sebagai berikut.

2.1 Emosi

Emosi dapat digambarkan sebagai keadaan yang pada umumnya disebabkan oleh suatu kejadian, seseorang, atau lingkungan luar lainnya atau biasa disebut dengan istilah stimulus. Pada saat manusia bereaksi terhadap lingkungannya maka emosi akan tercipta disaat bersamaan. Ketika emosi tercipta pada diri seseorang, pada umumnya manusia akan menunjukkan ekspresi wajah atau reaksi tubuh sesuai dengan emosi yang tercipta. Menurut ahli filosofi emosi, Rene Descartes, emosi terbagi menjadi enam yaitu heran, cinta, marah, hasrat, senang, dan sedih [1].

Terciptanya emosi berpengaruh pada kualitas kegiatan yang dilakukan manusia, seperti belajar. Emosi yang positif dapat mempercepat dan mempermudah proses belajar agar mendapatkan hasil yang lebih baik. Sebaliknya, emosi negatif dapat menghambat proses belajar bahkan hingga memberhentikan proses belajar manusia. Emosi negatif hanya akan menambahkan beban pada pikiran seseorang saat bekerja, sedangkan emosi positif membantu pikiran seseorang dengan munculnya ide-ide baru. Oleh karena itu, emosi positif perlu diciptakan untuk memulai pekerjaan yang baik [6].

2.2 Lirik Lagu

Lirik lagu merupakan bentuk ekspresi seseorang terhadap lingkungan dan pengalamannya yang dituliskan dalam kata-kata. Dalam mengekspresikan perasaannya, penyair atau pencipta lagu perlu melakukan permainan kata-kata dan bahasa untuk menciptakan daya tarik dan kekhasan terhadap lirik atau syairnya. Permainan bahasa ini dapat berupa permainan vokal, gaya bahasa maupun penyimpangan makna kata dan diperkuat dengan penggunaan melodi dan notasi musik yang disesuaikan dengan lirik lagunya sehingga pendengar semakin terbawa dengan apa yang dipikirkan pengarangnya [4].

2.3 Data Mining

Secara Bahasa, *data mining* adalah “menambang” atau mengekstraksi pengetahuan dari kumpulan data berukuran besar untuk mendapatkan informasi yang dapat dikembangkan. *Data mining* juga biasa disebut dengan istilah *Knowledge Discovery in Database* (KDD) yaitu proses transformasi data mentah menjadi informasi berguna. Beberapa ilmu yang berkaitan dengan *data mining* yaitu statistik, pembelajaran mesin, dan sistem basis data.

2.3.1 Preprocessing

Preprocessing adalah salah satu metode yang digunakan pada metode *data mining* yang bertujuan untuk mendapatkan pola data yang sama. Berikut ini beberapa tahapan pada *preprocessing*:

1) *Case Folding*

Tidak semua dokumen teks hanya menggunakan huruf kapital. Oleh karena itu, *Case folding* digunakan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf ‘a’ sampai dengan ‘z’ yang diambil, karakter selain huruf dihilangkan.

2) *Stemming*

Stemming merupakan suatu proses transformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya dengan aturan-aturan tertentu, atau dapat dianggap sebagai pembuangan imbuhan pada kata.

3) *Stopword Removal*

Tahapan ini adalah proses pengambilan kata-kata penting dan pembuangan kata-kata yang dianggap tidak penting. Proses pembuangan kata tidak penting dapat diterapkan pada *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). Contoh *stopword* adalah “yang”, “dan”, “di”, “dari”, dan sebagainya.

2.4 Part-of-Speech (POS) Tagging

Part-of-Speech (POS) *Tagging* yang juga disebut sebagai pelabelan kelas kata adalah suatu proses yang memberikan label kelas kata secara otomatis pada suatu kata dalam kalimat [9]. Proses yang dilakukan berupa melabelkan sebuah kata apakah itu kata kerja, kata sifat, atau hanya keterangan.

Tabel 2.1 Keterangan Label POS-Tagging

Label	Keterangan
CC (coordinate conjunction)	Konjungsi koordinasi
CD (cardinal numerals)	Bilangan kardinal
OD (ordinal number)	Bilangan urutan
FW (foreign words)	Kata serapan/kata asing
IN (prepositions)	Preposisi
JJ (adjectives)	Kata sifat
MD (modal or auxiliaries verbs)	Kata kerja bantu/modal
NEG (negations)	Negasi
NN (common nouns)	Kata benda umum, tidak spesifik
NNP (proper nouns)	Kata benda dalam huruf kapital
PR (common pronouns)	Pengolahan kata benda secara umum
PRP (personal pronouns)	Kata ganti orang
RB (adverbs)	Keterangan waktu
SC (subordinate conjunction)	Kata sambung/penghubung
SYM (symbols)	Simbol
VB (verbs)	Kata kerja

2.5 Pembobotan Kata (*Term Weighting*)

Proses pembobotan kata ini dilakukan untuk mendapatkan nilai dari kata (*term*). Pada penelitian ini akan digunakan metode *Term Frequency-Inversed Document Frequency* (*TF-IDF*) sebagai proses pembobotan, metode ini berfungsi untuk mencari representasi nilai dari masing-masing dokumen dari sekumpulan data

latih untuk dijadikan bentuk suatu vector dengan dokumen kata. Berikut persamaan yang digunakan untuk menghitung pembobotan kata [5].

$$W_{id} = tf \times IDF_t \quad (2.1)$$

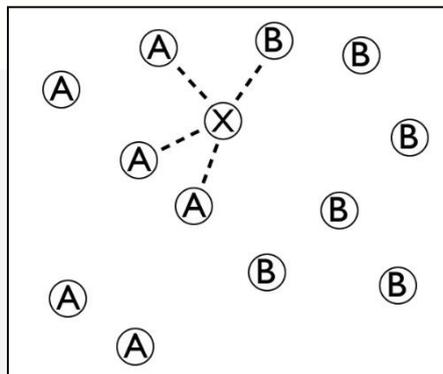
2.6 K-Nearest Neighbor (K-NN)

K-Nearest Neighbor (k-NN) adalah salah satu algoritma yang digunakan untuk melakukan klasifikasi. Algoritma k-NN termasuk metode *supervised learning* yang berarti adanya data latih dan terdapat variabel atau atribut yang ditargetkan sehingga tujuan dari algoritma ini adalah memetakan suatu data ke data yang sudah ada. Algoritma k-NN melakukan klasifikasi terhadap suatu objek berdasarkan data pembelajaran yang memiliki jarak terdekat dengan objek tersebut. Hal yang penting dalam algoritma ini adalah pemilihan nilai k, jika k sangat kecil maka akan menciptakan *noise*. Sebaliknya, jika terlalu besar dapat menciptakan N dengan banyak kelas yang harus diklasifikasikan.

Jika memasukan sebuah data *query* yang tidak diketahui labelnya, maka *K-Nearest Neighbor* (k-NN) akan mencari k buah data latih yang memiliki jarak terdekat dengan *query* dalam ruang n-dimensi. Pengukuran jarak antara data latih dengan *query* dilakukan dengan mengukur jarak antar titik yang merepresentasikan data *query* dengan semua titik yang merepresentasikan data latih menggunakan rumus *Euclidean Distance* berikut :

$$d(x, y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (2.2)$$

Dimana x dan y merupakan titik pada ruang vektor n dimensi sedangkan x_i dan y_i merupakan besaran skalar untuk dimensi ke i dalam ruang vektor n dimensi.



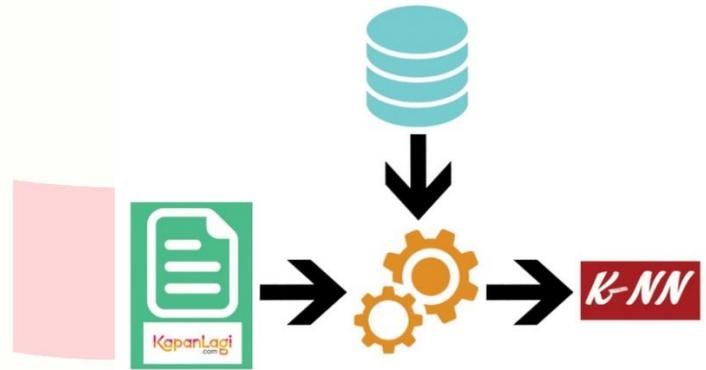
Gambar 2.1 Sebaran Data dalam Algoritma k-NN

Gambar 2.1 menunjukkan sebaran data dalam algoritma k-NN yang terdiri dari dua kelas yaitu kelas A dan B. Dalam hal ini, sebuah data uji akan diklasifikasikan berdasarkan data latih. Sebagai contoh, k yang digunakan adalah 4. Maka, data uji akan mencari 4 data latih yang berdekatan dengan data uji. 4 data tersebut adalah 1 titik menuju kelas B dan 3 titik menuju kelas A. Karena jumlah titik kelas A lebih dari titik kelas B, maka data uji akan diklasifikasikan sebagai kelas A

3. Perancangan

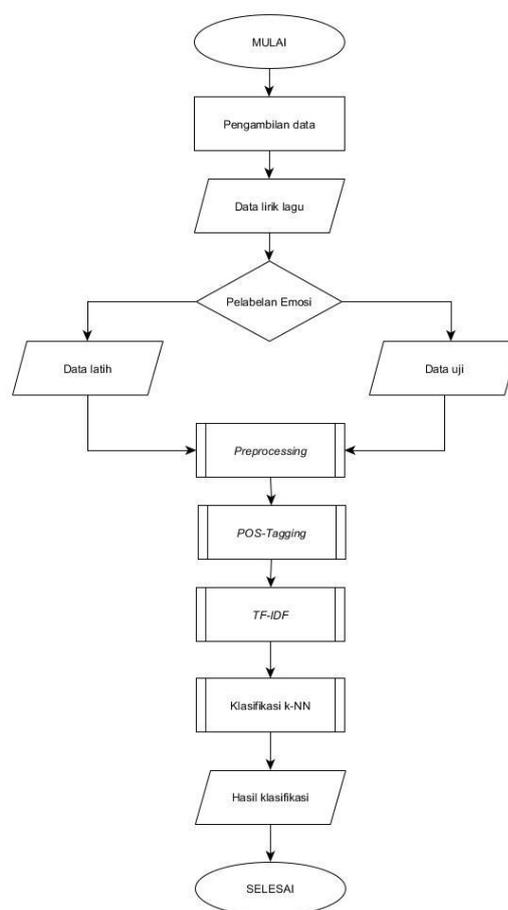
3.1 Perancangan Sistem Secara Umum

Sistem yang akan dibangun pada tugas akhir ini adalah sistem yang mampu mendeteksi kategori emosi pada lirik lagu berbahasa Indonesia dengan mengimplementasikan salah satu teknik *data mining*. Teknik yang digunakan adalah klasifikasi menggunakan algoritma *k-Nearest Neighbor*. *Study case* dalam tugas akhir ini adalah beberapa lirik lagu berbahasa Indonesia pada situs *lirik.kapanlagi.com*.



Gambar 3.1 Gambaran Umum Sistem

Gambar 3.1 diatas adalah gambaran umum sistem yang digunakan dalam tugas akhir ini. Data yang digunakan berupa lirik lagu berbahasa Indonesia diambil dari situs *lirik.kapanlagi.com*. Lirik lagu digunakan sebagai data latih maupun data uji nantinya.



Gambar 3.2 Flowchart Perancangan Sistem

Gambar 3.2 menjelaskan dimulai dari pengambilan data lirik lagu berbahasa Indonesia. Data latih dan data uji akan dilakukan pelabelan emosi secara manual. Pelabelan emosi dilakukan dengan cara mencocokkan kata-kata pada lirik lagu dengan *rules* berisi kata-kata yang mengandung label emosi. *Rules* yang dibuat sudah divalidasi oleh Balai Bahasa Jawa Barat. Tujuan pelabelan manual pada data latih adalah sebagai kelas yang digunakan untuk klasifikasi. Sedangkan tujuan pelabelan manual pada data uji bertujuan sebagai pengecekan kinerja sistem nantinya. Tahap selanjutnya adalah melakukan *preprocessing* untuk mendapatkan data yang bersih. Kemudian pembobotan TF-IDF, dan tahap terakhir adalah klasifikasi menggunakan algoritma *k-Nearest Neighbor*.

3.2 Klasifikasi k-Nearest Neighbor

Algoritma klasifikasi yang digunakan pada tugas akhir ini adalah *k-Nearest Neighbor*. Tujuan metode ini adalah untuk mengklasifikasikan objek ke dalam salah satu kelas terdekat yang telah ditetapkan dari kelas label emosi yang telah dibuat secara manual yaitu cinta, marah, senang dan sedih.

Masukan untuk proses k-NN adalah nilai yang didapat dari TF-IDF untuk dijadikan bentuk *vector space*. Berikut penjelasan lebih lanjutnya:

1. Hasil dari TF-IDF akan menjadi data latih yang sudah diberi label terhadap data uji. Bobot nilai TF-IDF akan membentuk *vector space model* untuk dijadikan variabel pada perhitungan *Euclidean Distance*.
2. Jarak antara data latih dan data uji dihitung dengan cara mengukur jarak antara titik yang mempresentasikan data latih dan data uji menggunakan rumus *Euclidean Distance*.

$$d = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad (3,1)$$

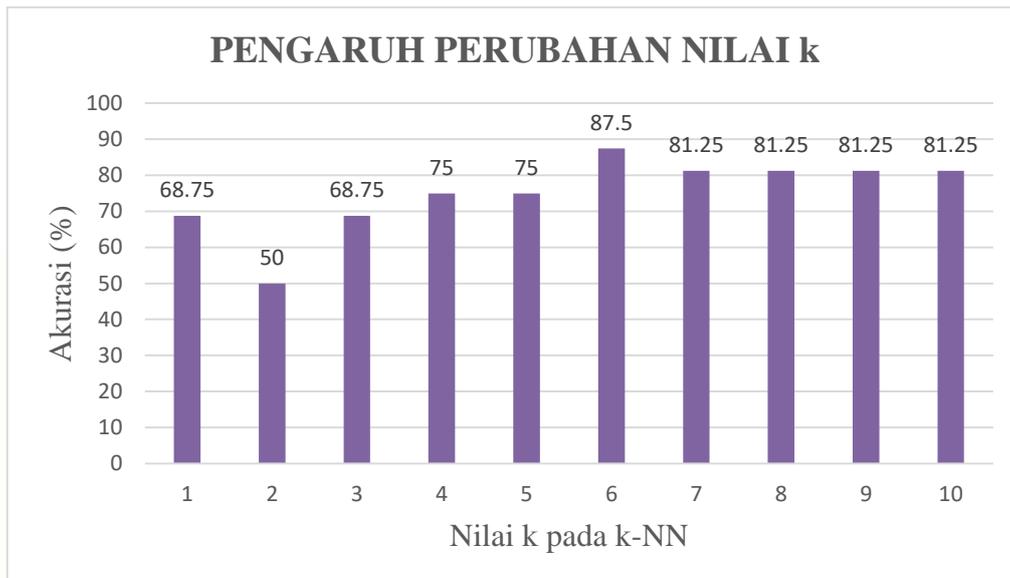
3. Mengurutkan data hasil perhitungan *Euclidean Distance*.
4. Jarak tersebut akan diurutkan dan dilakukan penentuan tetangga terdekat berdasarkan nilai k. Jumlah yang mayoritas akan digunakan sebagai hasil klasifikasi dari data latih.

4. Pengujian

Dari beberapa proses yang sudah dilakukan sebelumnya, selanjutnya dilakukan proses pengujian 16 data lirik lagu yang sudah dikumpulkan dan sudah diberi label emosi secara manual. Pengujian yang dilakukan adalah memasukkan seluruh data uji ke dalam sistem dan menghasilkan klasifikasi label emosi. Hal ini dilakukan bertujuan untuk menarik kesimpulan hasil akhir yang akan ditampilkan untuk mengetahui apakah program yang dirancang dapat mengklasifikasikan label emosi pada lirik lagu dengan benar.

4.1. Pengujian Nilai k pada k-NN Terhadap Akurasi Sistem

Pengujian mencari nilai k pada algoritma k-NN untuk mendapatkan akurasi terbaik pada sistem. Pengujian dilakukan menggunakan empat lirik lagu tiap label emosinya. Digunakan sebanyak 48 data latih lirik lagu dan 16 data uji lirik lagu (4 lagu/label emosi). Pada pengerjaan tugas akhir ini pengujian keberhasilan program dilihat dengan membandingkan *output* program dengan label emosi manual yang sudah divalidasi oleh Balai Bahasa Jawa Barat.



Gambar 4.1 Grafik Hasil Akurasi Sistem

Pada Gambar 4.1 ditampilkan grafik dari hasil pengujian sistem terhadap akurasi sistem. Terbukti bahwa nilai k pada algoritma k-NN akan mempengaruhi kinerja dari sistem yang telah dibuat. Terlihat grafik cenderung meningkat dari pengujian nilai k = 1 hingga k = 6 dan mencapai kondisi stagnan saat k = 7. Berdasarkan grafik diatas dapat diketahui nilai akurasi terbaik sistem didapat saat nilai k = 6 yaitu sebesar 87,5%. Dan didapatkan hasil rata-rata akurasi dari pengujian 10 kali sebesar 75

5. Kesimpulan

Berdasarkan dari hasil Tugas Akhir ini, dapat ditarik beberapa kesimpulan yaitu:

1. Proses klasifikasi data lirik lagu hanya diberikan 4 jenis kelas label emosi yaitu cinta, marah, senang, dan sedih menggunakan algoritma k-*Nearest Neighbor*.
2. Sistem yang dirancang berhasil mengklasifikasikan emosi berdasarkan lirik lagu secara otomatis.
3. Nilai k terbaik adalah k = 6 dengan akurasi sebesar 87,5% dan rata-rata pada 10 kali pengujian memiliki nilai akurasi sebesar 75%.

Daftar Pustaka:

- [1] H. David, *Emotion And Moods*, pp. 259-297.
- [2] S. Swathi and G. S. E., "Current Emotion Research in Music Psychology," *Emotion Review*, vol. 7, no. 2, pp. 189-197, 2015.
- [3] Y. Dan and L. Won-Sok, "Music Emotion Identification from Lyrics," in *IEEE International Symposium on Multimedia*, San Diego, 2009.
- [4] S. Satria, "Eksperimen Mengenai Emosi Oleh Musik Bahagia Dan Sedih Dengan Volume Terentu," 2010.
- [5] S. Lailatus, A. Zainal and N. Hani, "Klasifikasi Emosi Untuk Teks Berbahasa Indonesia Dengan Menggunakan K-Nearest Neighbor," in *National Conference Green Technology 3*, Malang, 2012.
- [6] D. R. Anna and F. Julie, "Understanding the Role of Negative Emotions in Adult Learning and Achievement: A Social Functional Perspective," *Behav. Sci.*, vol. 8, no. 27, 2018.
- [7] H. Suhartina, N. Muhammad, S. Casi and A. M. Dr. Muhammad, "Classification Of Posts Twitter Traffic Jam The City Of Jakarta Using Algorithm C4.5," in *International Conference on Signals and Systems (ICSigSys)*, Bali, 2018.
- [8] Y. H. Romario, R. M. P. Rekyan and C. W. Randy, "Implementasi Metode K-Nearest Neighbour Dengan Pembobotan TF.IDF.ICF Untuk Kategorisasi Ide Kreatif Pada Perusahaan," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 4, no. 2, pp. 97-103, 2017.
- [9] D. Arawinda, R. Fam, L. Andry and M. Ruli, "Designing an Indonesian Part of speech Tagset and Manually Tagged Indonesian Corpus".
- [10] G. Gongde, W. Hui, B. David, B. Yaxin and G. Kieran, "KNN Model-Based Approach in Classification".