SCIENCE AND INFORMATION TECHNOLOGY
SINTECH
JOURNAL

SINTECH JOURNAL | ISSN 2598-7305 | E-ISSN 2598-9642
Vol. 7 No 1 – April 2024 | https://s.id/sintechjournal
DOI: https://doi.org/10.31598
Publishing: Prahasta Publisher

Detecting Emotions of Indonesian Songs Based on Plutchik's Theory using Data Mining

Deyana Kusuma Wardani[1*], Dwi Diana Wazaumi[2], Raden Rara Kartika Kusuma Winahyu [3]

[12]Informatics Management, Astra Polytechnic,
Jl. Gaharu Blok F-3 Delta Silicon 2 Lippo Cikarang, Kel. Cibatu, Kec. Cikarang Selatan, Bekasi, Jawa Barat

e-mail: deyana.kusuma@polytechnic.astra.ac.id[1], dwi.diana@polytechnic.astra.ac.id[2],
raden.rara@polytechnic.astra.ac.id[3]

Received : January, 2024	Accepted : February, 2024	Published : April, 2024

Abstract

Listening to songs is a daily activity that everyone engages in. Most people choose songs based on their mood, so a system is needed to detect emotions from song lyrics. Previous research only focused on five basic emotions: happy, sad, love, anger, and fear. In this study, we propose a new method to detect emotions from song lyrics using Plutchik's emotion theory. According to this theory, we can detect more emotions from song lyrics. The data generated consists of 300 data, but this research utilizes only 250 of them as it focuses solely on selecting the top Indonesian songs. This research categorizes human emotions into eight: joy, trust, surprise, sadness, disgust, anger, and anticipation. Next, calculate the threshold value using the mean formula: divide the sum of all frequencies of each emotion by the number of emotions in Plutchik's model. This value is used to determine the dominant emotion. If the frequency value of an emotion is higher than the threshold value, the system considers it as the dominant emotion. The dominant emotions are then classified into positive and negative emotions using cosine similarity calculations. The sampling technique uses 70% training data and 30% testing data. The accuracy obtained using the classification report is 0.81, indicating that the method has a good level of accuracy.

Keywords: *plutchik's theory, data preprocessing, data mining, cosine similarity*

Abstrak

Mendengarkan lagu adalah kegiatan sehari-hari yang dilakukan semua orang. Kebanyakan orang memilih lagu berdasarkan suasana hati mereka, sehingga diperlukan sebuah sistem untuk mendeteksi emosi dari lirik lagu. Penelitian sebelumnya hanya berfokus pada lima emosi dasar: senang, sedih, cinta, marah, dan takut. Pada penelitian ini, kami mengusulkan metode baru untuk mendeteksi emosi dari lirik lagu dengan menggunakan teori emosi Plutchik. Menurut teori ini, kita dapat mendeteksi lebih banyak emosi dari lirik lagu. Data yang dihasilkan terdiri dari 300 data, tetapi penelitian ini hanya menggunakan 250 data karena hanya berfokus pada pemilihan lagu-lagu Indonesia teratas. Penelitian ini mengkategorikan emosi manusia menjadi delapan, yaitu: kegembiraan, kepercayaan, keterkejutan, kesedihan, rasa jijik, kemarahan, dan antisipasi. Selanjutnya, hitung nilai ambang batas dengan menggunakan rumus rata-rata: bagi jumlah semua frekuensi setiap emosi dengan jumlah emosi dalam model Plutchik. Nilai ini digunakan untuk menentukan emosi yang dominan. Jika nilai frekuensi suatu emosi lebih tinggi dari nilai ambang batas, sistem akan menganggapnya sebagai emosi dominan. Emosi dominan kemudian

diklasifikasikan menjadi emosi positif dan negatif dengan menggunakan perhitungan kemiripan kosinus. Teknik pengambilan sampel menggunakan 70% data latih dan 30% data uji. Akurasi yang diperoleh dengan menggunakan laporan klasifikasi adalah sebesar 0,81 yang menunjukkan bahwa metode tersebut mempunyai tingkat akurasi yang baik.

Kata Kunci: *teori plutchik, data preprocessing, data mining, cosine similarity*

1. INTRODUCTION

Listening to songs has become a daily activity for everyone. In 2023, Spotify recorded that the number of people listening to music reached 547.000.000 [1]. The purpose of listening to songs is reduces depression or stress, raise the mood, and improve your health [2]. Songs have collaboration of rhythm and lyrics. Every rhythm and lyric created by a songwriter aims to send a message and emotions contained in the song.

Rhythm is a regular basic flow following variations of melodic movement [3]. Habitual listening to songs can make someone feel every emotion in the song. Difference in rhythm, which uses melody to convey a emotions of a song, lyrics use words that contain a message or emotions both implicit and explicit [4]. Songs lyrics from inspiration and life experience of songwriter [5].

Nowadays, most people listen to music according to their mood, so there is a need for a system that can detect emotions from song lyrics to make it easier to search for songs based on mood. There have been several studies [6]–[10] that can detect emotions from song lyrics, each using different methods. For example, Research [6] utilizes LSTM and Word2Vec, while Research [7] employs a multi-class support vector machine with particle swarm optimization. Furthermore, Research [8] utilizes support vector machine and particle swarm optimization, and Research [9] uses the naive Bayes algorithm and particle swarm optimization. In contrast, Research [10] takes a descriptive qualitative approach, unlike the previous four studies [6]–[9].

The five previous studies [6]–[10] have in common that they only used these five basic emotions: happy, sad, love, anger, and fear; whereas basic human emotions can be more than five. Plutchik's theory states that basic human emotions are divided into eight: pleasure, trust, fear, surprise, sadness, disgust, anger, anticipation [11]. This research proposes a new method, namely emotion classification using Plutchik's theory of emotions from top Indonesian songs. The purpose of using this theory is to detect more emotions in song lyrics because this theory can detect eight basic human emotions, which is more than the emotions that can be detected in previous studies.

Basically, emotions are categorized into positive emotions and negative emotions as well as Plutchik's emotions [12]. Plutchik's eight emotions are grouped into positive emotions and negative emotions using the cosine similarity method. The data used is in the form of text, previous research [13] used this method to find content similarity from text data. The result is that this method is effectively used to calculate the similarity of two text documents. From this we use the cosine similarity method to calculate the similarity of texts with lyrics and words containing emotions.

2. RESEARCH METHOD

The proposed research introduces a novel approach to classify emotions in Indonesian songs. Emotion theory by Plutchik and the cosine similarity method are utilized for this classification. The process flow, starting from data collection to the classification process, is depicted in Figure 1.

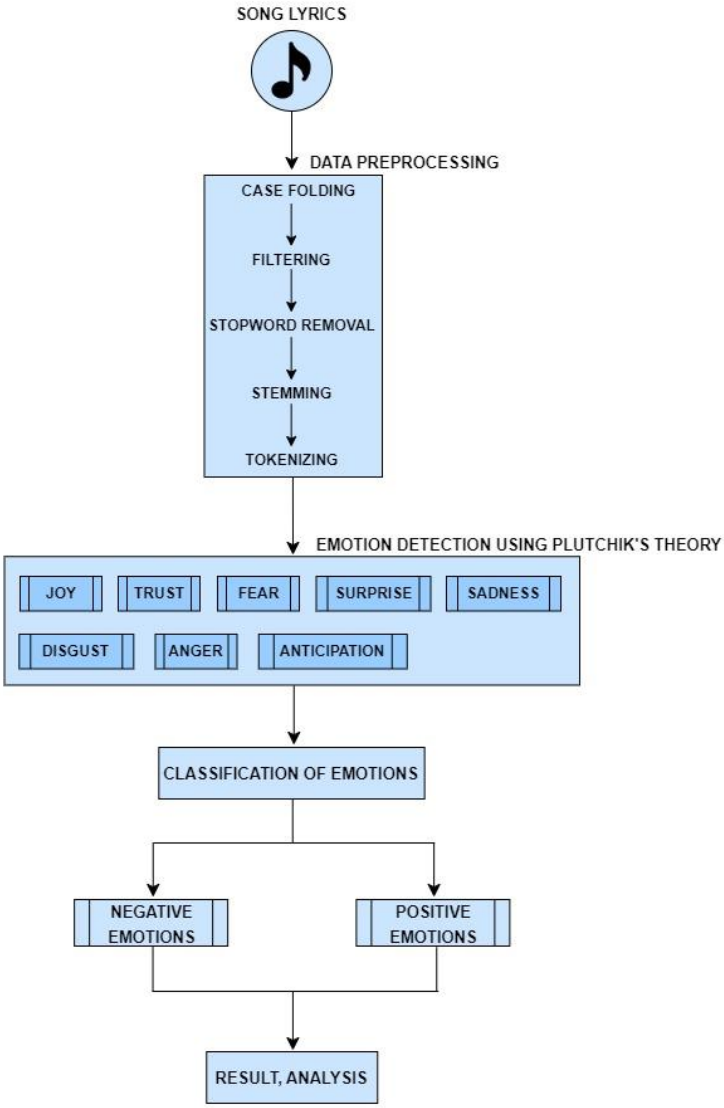

Figure 1. System Design

The data was obtained from Shazam.com, a mobile application that utilizes audio recognition technology to identify music, movies, TV shows, and other media [14]. To retrieve the data from this website, we utilized a Laravel package called Spatie. We specifically selected the top Indonesian songs from Spotify, Youtube, and Apple Music. In total, the dataset comprises 250 song lyrics, with 30% of the data allocated for testing purposes.

The data still requires cleaning before the classification process. To achieve this, we performed data preprocessing. The initial step involved case folding, where all uppercase alphabets were converted to lowercase [15].

Table 1. Case Folding Process

Sentence	Case Folding
Bersamamu kita akan taklukan dunia. Arungi samudra penuh rintangan.	bersamamu kita akan taklukan dunia. arungi samudra penuh rintangan.

The second stage in the data preprocessing is filtering. Once all the sentences have been converted to lowercase, we proceed to remove numbers, symbols, punctuation marks, and links [16]. Table 2 provides an illustration of the filtering process.

Table 2. Filtering Process

Sentence	Filtering
bersamamu kita akan taklukan dunia. arungi samudra penuh rintangan.	bersamamu kita akan taklukan dunia arungi samudra penuh rintangan

Each sentence in the data contains words that are considered stopwords, which do not carry significant meaning [17]. To enhance the data quality, it is necessary to eliminate these stopwords. Since the research data is in Indonesian, we utilized the Sastrawi package from Python [18] for the stopword removal process.

Table 3. Stopword Removal Process

Sentence	Stopword Removal
bersamamu kita akan taklukan dunia arungi samudra penuh rintangan	bersamamu akan taklukan dunia arungi samudra penuh rintangan

The fourth stage involves stemming, where each word is transformed into its base form [19], eliminating any affixes. Similar to the stopword removal stage, we utilized the Sastrawi package for this process. Table 4 provides an illustration of this stage.

Table 4. Stemming Process

Sentence	Stemming
bersamamu akan taklukan dunia arungi samudra penuh rintangan	sama akan takluk dunia arung samudra penuh rintang

After all processes are done, the next stage is trimming the sentence into words [20] and changing them into tokens using the tokenizing process. The purpose of this stage is make it easier for classification process.

Table 5. Tokenizing Process

Sentence	Tokenizing
sama akan takluk dunia arung samudra penuh rintang	[['sama','akan','takluk','dunia','arung', 'samudra','penuh','rintang']]

Robert Plutchik divided emotions into 8 (eight) categories, which are joy, trust, fear, surprise, sadness, disgust, anger, and anticipation [11]. Among these eight emotions, they will be classified into positive emotions and negative emotions.

To detect emotions from song lyrics, we used the NRC Word-Emotion Association Lexicon by Saif Mohammad [21]. This dictionary contains a group of words from various language and classified based on plutchik's emotions theory. Due to the utilization of Indonesian data, we also used this dictionary in the Indonesian language.

Following preprocessing stage, data was matched with each emotion in NRC Word-Emotion Association Lexicon to determine the frequency value. Next frequency value of each emotion will be calculated using the sum() function in Python. An ilustration of this process is shows in table 6.

Table 6. Ilustration Anger Emotion

Words	Frequency
Ditinggalkan	0

Benci	2
Mengerikan	0
Menghapuskan	0
Kekejian	0
Melampiaskan	7
Memeras	0
Menyalak	0
Terkutuk	1
Tuduhan	0
Total	10

Afterwards, we established a threshold value to identify the dominant emotions among the eight Plutchik's emotions. We utilized formula 1 is mean formula by adding up the frequency values of each emotion and dividing them by eight, which represents the total number of Plutchik's emotions.

$$\bar{x} = \frac{\sum x}{n} \quad (1)$$

$\bar{x}$ is mean value, $\sum x$ is total all frequency value of each emotion, and n is total number of plutchik's emotions.

Next, we implemented conditions where if the frequency value of each emotion is greater than the threshold, we assigned a value of 1 (one), indicating it as a dominant emotion. Otherwise, we assigned a value of 0 (zero) to indicate it as non-dominant.

The eight emotions from plutchik's theory can be classified into two categories : four positive emotions and four negative emotions. Four positive emotions are joy, trust, fear, and surprise. These emotions have corresponding opposite emotions, which are sadness, disgust, anger, and anticipation [12].

We created an emotions dictionary that includes both positive and negative emotions. If a particular emotion is classified as positive or negative, we assigned it a value of 1; otherwise, we assigned a value of 0. Then, we proceeded to create an emotions matrix.

Table 7. Emotions Dictionary with Emotions Matrix

emotion categories	**plutchik's emotions**							
	joy	trust	fear	surprise	sadness	disgust	anger	anticipation
positive	1	1	1	1	0	0	0	0
negative	0	0	0	0	1	1	1	1

To detect emotions, we utilized formula 2, which is cosine similarity [13]. Cosine similarity is an algorithm used to calculate the similarity between two documents by comparing their values and determining the similarity score.

$$similarity = \frac{\bar{x} \cdot \bar{y}}{||\bar{x}|| \cdot ||\bar{y}||}$$

Similarity is similarity score between emotions dictionary. $\bar{x}$ is dominant emotions and $\bar{y}$ is emotions dictionary.

Next, using formula 2, we compared the dominant emotions with the emotions dictionary. Then, we obtained two similarity values, one for the positive category and one for the negative category. We used the max() function in Python to select the higher similarity value between the two categories. Finally, we assigned the label of either positive or negative emotions to the song lyrics.

3. RESULT AND PERFORMANCE MEASUREMENT

3.1 Result

This research divided emotions into two categories : positive emotions and negative emotions, based on plutchik's theory. All the data from song lyrics have already been detected for emotions and table 8 illustrates the results of the emotion detection process.

Table 8. Illustration result

Song Lyrics	System
… Dia buatku nyaman. Dalam hangat pelukan. Dia perasa. Yang mengerti yang kurasa. Hanya dia…	Positive
Bersamamu kita akan taklukan dunia. Arungi samudra penuh rintangan. Bersamamu kita akan jalani semua. Jangan takut untuk melangkah bersamaku…	Positive
Dari kejauhan tergambar cerita tentang kita. Terpisah jarak dan waktu. Ingin kuungkapkan rindu lewat kata indah. Tak cukup untuk dirimu…	Positive
Sekian lamanya kita tak berjumpa. Namun, tak sengaja berpapasan ku denganmu. Tatapan yang sama, rasanya yang lama. Namun, ku kecewa, kau tak peka…	Negative
Denting yang berbunyi dari dinding kamarku. Sedarkan diriku dari lamunan panjang. Tak terasa malam kini semakin larut. Ku masih terjaga…	Negative

3.2 Performance Measurement

The accuracy calculation is utilized to assess the effectiveness of the employed method. The total number of data is 250. We apply a sampling technique, 70% training data and 30% testing data, so the testing data used is 75 data. Next, we manually assigned labels to the testing data. The manual labels were then compared with the results generated by the system. For the labels, we replaced positive emotions with the number 1 and negative emotions with -1.

We utilized the Python library, Scikit-Learn, to calculate the accuracy of the method. From this library, we employed two objects: confusion_matrix and classification_report. The confusion_matrix object was used to assess accuracy by computing the confusion matrix, resulting in Figure 6.

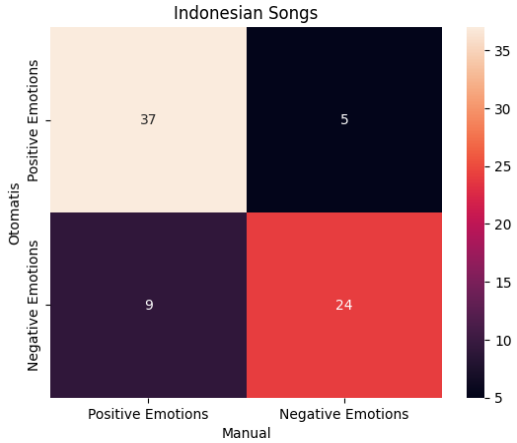

Figure 2. The outcome of Confusion Matrix

Figure 6 has two classes: "automatic" represents the predicted class, while "manual" represents the actual class. The TP (True Positive) value is colored pink with a value of 37; the FP (False Positive) value is colored black with a value of 5; the FN (False Negative) value is colored purple with a value of 9; and the TN (True Negative) value is colored red with a value of 24.

Then we used classification_report to evaluate the performance of the method with a detailed classification report. This report includes precision, recall, f1-score, support, and accuracy. Table 9 displays the classification report of this research.

Table 9. Classification Report

Classification Report :				
	precision	recall	f1-score	support
1	0.80	0.88	0.84	42
-1	0.83	0.73	0.77	33
accuracy			0.81	75
macro avg	0.82	0.80	0.81	75
weighted avg	0.81	0.81	0.81	75

The accuracy of the method is 0.81. The data bias occurs because the frequency value of the emotions in the song is not greater than the threshold value, thereby preventing them from being considered as dominant emotions.

4. CONCLUSION

This research proposes a new method to detect emotions from song lyrics. The method used is to detect emotions using Plutchik's theory, which divides the basic emotions of the song into eight : joy, trust, fear, surprise, sadness, disgust, anger, and anticipation. We use thresholds to determine the dominant emotion among the eight emotions. We classify positive emotions and negative emotions based on the opposite emotions from Plutchik's theory. The cosine similarity method is used to classify the lyrics into positive emotions or negative emotions.

This research has successfully classified emotions from song lyrics based on Plutchik's theory with an accuracy of 0.81 using 30% of the testing data. Song lyrics that do not match the intended emotion are due to the frequency value of the song's emotion not exceeding the threshold value. This research can be used to automatically classify emotions from song lyrics on several platforms for listening to songs such as Spotify, Youtube, and Apple Music.

REFERENCES

[1] Spotify, "The Top Songs, Artists, Podcasts, and Listening Trends of 2023 Revealed," 2023. https://newsroom.spotify.com/2023-11-29/top-songs-artists-podcasts-albums-trends-2023/#:~:text=As always%2C our 2023 Spotify,the world listened this year

[2] F. S. S. Sinaga and E. Winangsit, "Terapi Musik untuk Meningkatkan Kesehatan Mental: Tinjauan Literatur dalam Perspektif Psikodinamika," *Assert. Islam. Couns. J.*, vol. 02, no. 1, pp. 1–12, 2023.

[3] A. M. W. Aru, "E-Modul Seni Budaya Kelas XI," *Angew. Chemie Int. Ed. 6(11), 951–952.*, pp. 5–24, 2018.

[4] A. Yanti, P. A. Lubis, N. A. Natasha, E. Sitorus, and F. L. Barus, "Analisis Makna Leksikal Pada Lirik Lagu Kamu Dan Kenangan Karya Maudy Ayunda," *Asas J. Sastra*, vol. 2, no. 3, p. 86, 2021, doi: 10.24114/ajs.v10i2.26269.

[5] F. R. Zain, "Memahami Hakikat Kehidupan Melalui Lirik Lagu Karya Banda Neira (Tinjauan Semantik: Metafora)," *Kaji. Linguist. dan Sastra*, vol. 6, no. 2, pp. 164–175, 2021, doi: 10.23917/kls.v6i2.9589.

[6] I. P. D. Fortunawan, N. Agus, and S. Er, "Klasifikasi Emosi Lirik Lagu dengan Long Short-Term Memory dan Word2Vec," vol. 1, pp. 1203–1208, 2023.

[7] H. Sastypratiwi, H. Muhardi, and M. Noveanto, "Klasifikasi Emosi Pada Lirik Lagu Menggunakan Algoritma Multiclass SVM dengan Tuning Hyperparameter PSO," *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2279, 2022, doi: 10.30865/mib.v6i4.4609.

[8] W. Hermanto, B. Irawan, C. Setianingsih, and U. Telkom, "Klasifikasi Emosi Pada Lirik Lagu Menggunakan Algoritma Support Vector Machine Dan Optimasi Particle Swarm Optimization

Classification of Emotions in Song Lyrics Using Support Vector Machine Algorithm and Particle Swarm," vol. 8, no. 5, pp. 6307–6327, 2021.

[9] G. S. Ramadhan, B. Irawan, C. Setianingsih, and U. Telkom, "Klasifikasi emosi pada lirik lagu menngunakan algoritma naïve bayes dan particle swarm optimization classification of emotions on song lyrics using naïve bayes algorithm and particle swarm optimization," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 6293–6306, 2021.

[10] R. Damara, R. Kezia, G. Bagus, and H. Putri, "IJM : Indonesian Journal of Multidisciplinary Klasifikasi Emosi yang Terdapat Dalam Lirik Lagu ' Jiwa yang Bersedih ' Karya Ghea Indrawari," vol. 1, pp. 2139–2147, 2023.

[11] W. Alan, "Klasifikasi Emosi pada teks menggunakan Deep Learning," *Tugas Akhir di Univ. Islam Indones.*, vol. 6, no. 1, 2021.

[12] A. Mondal and S. S. Gokhale, "Mining Emotions on Plutchik's Wheel," *2020 7th Int. Conf. Soc. Netw. Anal. Manag. Secur. SNAMS 2020*, no. December 2020, 2020, doi: 10.1109/SNAMS52053.2020.9336534.

[13] F. A. Nugroho, F. Septian, D. A. Pungkastyo, and J. Riyanto, "Penerapan Algoritma Cosine Similarity untuk Deteksi Kesamaan Konten pada Sistem Informasi Penelitian dan Pengabdian Kepada Masyarakat," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 529, 2021, doi: 10.32493/informatika.v5i4.7126.

[14] Shazam, "Shazam," 2023. https://www.shazam.com/

[15] R. A. Sihombing and F. Natsir, "Perancangan Aplikasi Analisis Sentimen Terhadap Opini Penghapusan Skripsi pada Twitter menggunakan Metode Naïve Bayes," vol. 2, no. 1, pp. 1–14, 2023.

[16] Deyana Kusuma Wardani, Iwan Syarif, and Tessy Badriyah, "Detecting Alter Ego Accounts using Social Media Mining," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 3, pp. 602–608, 2023, doi: 10.29207/resti.v7i3.4919.

[17] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *JST (Jurnal Sains dan Teknol.*, vol. 12, no. 1, pp. 129–137, 2023, doi: 10.23887/jstundiksha.v12i1.52358.

[18] A. E. Budiman and A. Widjaja, "Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 3, pp. 475–488, 2020, doi: 10.28932/jutisi.v6i3.2892.

[19] M. Afdal and L. R. Elita, "Penerapan Text Mining Pada Aplikasi Tokopedia Menggunakan Algoritma K-Nearest Neighbor," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 8, no. 1, p. 78, 2022, doi: 10.24014/rmsi.v8i1.16595.

[20] G. R. Yunarfi, Ricky Simdy, and Jackson, "Implementasi Text Mining untuk Mengetahui Kata Abreviasi dalam Percakapan Media Sosial," *J. Digit. Ecosyst. Nat. Sustain.*, vol. 1, no. 2, pp. 78–83, 2021.

[21] S. and Mohammad and P. Turney, "NRC Publications Archive Archives des publications du CNRC," 2002, doi: https://doi.org/10.4224/21270984.