

Analisis Emosi dalam Lirik Lagu menggunakan *Natural Language Processing*

Michael Sabda Husada ^{1*}, Sri Yulianto J.P ²

^{1,2}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Kota Salatiga, Provinsi Jawa Tengah, Indonesia.

Email: 672021078@student.uksw.edu ^{1*}, sri.yulianto@uksw.edu ²

Histori Artikel:

Dikirim 20 Oktober 2025; *Diterima dalam bentuk revisi* 25 Oktober 2025; *Diterima* 20 Desember 2025; *Diterbitkan* 10 Januari 2026. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Musik merupakan medium universal untuk mengekspresikan emosi, dengan lirik lagu sebagai komponen naratif yang kaya akan muatan perasaan. Penelitian ini bertujuan untuk menganalisis lanskap emosi dalam lirik lagu populer berbahasa Inggris yang dikumpulkan dari platform Spotify dan menguji efektivitas pendekatan Natural Language Processing (NLP) dalam mengklasifikasikannya. Korpus penelitian terdiri dari 57.494 lirik lagu yang dikumpulkan secara acak tanpa batasan genre. Melalui pipeline analisis yang komprehensif mulai dari pra pemrosesan teks (case folding, normalisasi, cleaning, tokenisasi, filtering, stemming), pelabelan berbasis leksikon emosi sendiri (lexicon-based labelling), ekstraksi fitur TF-IDF, hingga klasifikasi dengan model Random Forest penelitian ini berhasil mengungkap dua temuan utama. Secara empiris, lirik lagu didominasi oleh emosi positif, dengan 'romantic' (36,2%) dan 'happy' (26,2%) sebagai arus utama, diikuti oleh 'sad' (16,3%), sementara ekspresi 'angry' (4,4%) paling jarang ditemui, menunjukkan class imbalance yang signifikan. Secara metodologis, model yang dibangun menunjukkan kinerja yang solid dengan akurasi keseluruhan 83,03% dan *weighted avg F1-score* 0,82. Namun, analisis confusion matrix dan classification report mengungkap ketimpangan kinerja model antar kelas, di mana emosi 'angry' dan 'energetic' memiliki recall yang rendah (masing-masing 42% dan 62%), diduga kuat akibat ketidakseimbangan distribusi data dan keterbatasan leksikon dalam menangkap konteks. Kesimpulannya, penelitian ini tidak hanya berhasil memetakan dominasi tema cinta dan kebahagiaan dalam lirik lagu populer, tetapi juga mendemonstrasikan bahwa pendekatan NLP dengan model klasik dapat mencapai kinerja yang kompetitif. Temuan sekaligus menyoroti pentingnya penanganan class imbalance dan pengembangan leksikon yang lebih kaya konteks atau penggunaan model berbasis deep learning untuk penelitian di masa depan, guna menangkap spektrum emosi lirik secara lebih merata dan holistik.

Kata Kunci: Analisis Emosi; Lirik Lagu; Natural Language Processing; Random Forest; TF-IDF; Lexicon-based Labelling; Class Imbalance; Emotion Taxonomy.

Abstract

Music is a universal medium for expressing emotions, with song lyrics serving as a narrative component rich in affective content. This study aims to analyze the emotional landscape within popular English song lyrics collected from the Spotify platform and to examine the effectiveness of Natural Language Processing (NLP) approaches in classifying these emotions. The research corpus consists of 57,494 randomly collected song lyrics without genre restrictions. Through a comprehensive analytical pipeline---ranging from text preprocessing (case folding, normalization, cleaning, tokenization, filtering, stemming), custom lexicon-based emotion labeling, TF-IDF feature extraction, to classification using a Random Forest model---the study reveals two key findings. Empirically, song lyrics are dominated by positive emotions, with romantic (36.2%) and happy (26.2%) emerging as the main themes, followed by sad (16.3%), while angry expressions (4.4%) appear least frequently, indicating significant class imbalance. Methodologically, the proposed model demonstrates solid performance with an overall accuracy of 83.03% and a weighted avg F1 score of 0.82. However, analysis of the confusion matrix and classification report uncovers performance disparities across emotion classes: angry and energetic emotions exhibit low recall (42% and 62%, respectively), likely due to imbalanced data distribution and lexicon limitations in capturing context. In conclusion, this study not only succeeds in mapping the dominance of love- and happiness-related themes in popular song lyrics but also demonstrates that classical NLP models can achieve competitive performance. The findings additionally highlight the importance of addressing class imbalance and developing more context-rich emotion lexicons or employing deep learning models in future research, in order to capture the emotional spectrum of lyrics more evenly and comprehensively.

Keyword: Emotional Analysis; Song Lyrics; Natural Language Processing; Random Forest; TF-IDF; Lexicon-based Labelling; Class Imbalance; Emotion Taxonomy.

1. Pendahuluan

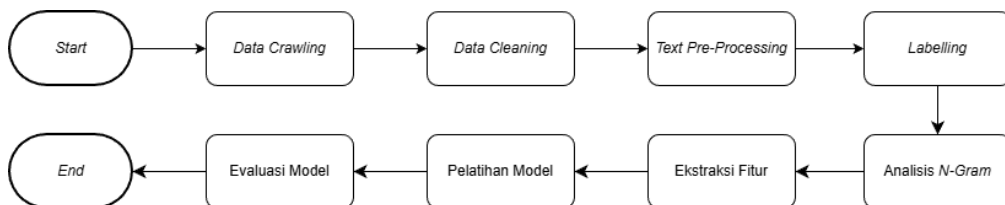
Musik diakui sebagai bentuk ekspresi manusia yang bersifat universal, memiliki kekuatan untuk menyampaikan perasaan melalui melodi dan lirik. Lirik lagu bukan hanya alat hiburan, melainkan juga medium untuk menyampaikan cerita dan emosi yang kompleks mulai dari kebahagiaan, cinta, hingga kesedihan dan kemarahan (Ardianti *et al.*, 2023; Gayatri & Zulfiningrum, 2025; Hutabarat *et al.*, 2025; Simbolon *et al.*, 2025). Pemahaman terhadap emosi dalam lirik lagu memberikan manfaat lebih dari sekadar perspektif sosiologis atau psikologis. Aplikasi praktisnya sangat luas, terutama di bidang rekomendasi musik yang disesuaikan dengan preferensi pengguna, pemantauan kesehatan mental, serta analisis perkembangan budaya (Chowdary *et al.*, 2024; Jo & Kim, 2022). Namun, tantangan utama terletak pada upaya untuk mengkuantifikasi dan mengklasifikasikan emosi yang terkandung dalam lirik lagu. Karakter subjektif dan penggunaan metafora yang kerap muncul, bersama dengan gaya bahasa tidak formal atau slang, menjadikan analisis manual tidak efisien dan sulit untuk diskalakan. Natural Language Processing (NLP) menawarkan solusi yang memungkinkan pemrosesan otomatis atas teks lirik dengan mengubahnya menjadi representasi numerik yang bisa dianalisis oleh model pembelajaran mesin untuk mengidentifikasi pola emosional (Ara & Rekha, 2024). Perkembangan teknologi dalam bidang ini menunjukkan berbagai pendekatan yang semakin kompleks. Ara & Rekha (2024) berhasil meningkatkan akurasi klasifikasi emosi dengan menggabungkan berbagai fitur NLP yang memanfaatkan leksikon emosi. Bužić & Dobša (2025) menggunakan NRC Emotion Lexicon untuk memetakan perubahan emosi dalam lirik kontes Eurovision, yang mengungkapkan pergeseran antara emosi positif dan negatif. Sementara itu, penggunaan model neural network yang lebih dalam, seperti Bidirectional LSTM (BiLSTM) dengan embedding kata GloVe oleh Abdillah *et al.* (2021), menunjukkan hasil yang menjanjikan dengan tingkat akurasi mencapai 91,08%. Di sisi lain, penelitian oleh Chowdary *et al.* (2024) menghubungkan emosi dalam lirik lagu dengan indikator kesehatan mental, memperluas aplikasi NLP dalam konteks yang lebih luas daripada sekadar hiburan.

Meski demikian, ada beberapa kekurangan yang masih perlu diperhatikan. Sebagian besar penelitian fokus pada penggunaan model-model kompleks, sementara efektivitas model klasik seperti Random Forest dalam menangani data besar masih minim dibahas. Selain itu, ketidakseimbangan kelas dalam dataset lirik lagu populer di mana tema cinta dan kebahagiaan mendominasi menyebabkan kesulitan dalam mengenali emosi yang jarang muncul, seperti kemarahan (Jo & Kim, 2022). Leksikon standar seperti NRC (Mohammad, 2023) juga belum sepenuhnya mampu menangkap nuansa kontekstual dalam lirik, terutama pada kategori yang tidak terdefinisi dengan jelas, seperti 'energetic' atau 'romantic' (Akaishi *et al.*, 2022; Jim *et al.*, 2024). Berdasarkan kekurangan-kekurangan tersebut, penelitian ini bertujuan untuk memberikan jawaban atas beberapa pertanyaan utama: (1) Bagaimana emosi terdistribusi dalam lirik lagu populer berbahasa Inggris? (2) Seberapa efektif model Random Forest dengan fitur TF-IDF dan leksikon kustom dalam mengklasifikasikan emosi? (3) Bagaimana ketidakseimbangan distribusi data memengaruhi kinerja model, terutama dalam mendeteksi emosi yang jarang muncul? Penelitian ini akan membangun pipeline analisis dimulai dari pengumpulan data besar dari Spotify, diikuti dengan serangkaian proses pra-pemrosesan teks (seperti case folding, normalisasi, pembersihan, tokenisasi, penyaringan, dan stemming), serta pelabelan emosi menggunakan leksikon kustom. Model Random Forest akan digunakan karena kemampuannya dalam menangani data berdimensi tinggi dan ketahanannya terhadap overfitting (Ngo *et al.*, 2022). Harapan dari penelitian ini adalah menghasilkan sebuah kerangka kerja yang dapat diandalkan, yang juga dapat diterapkan untuk menganalisis berbagai perasaan yang ada dalam lirik lagu.

2. Metode Penelitian

Penelitian ini bertujuan untuk menganalisis emosi dalam lirik lagu dengan menerapkan pendekatan Natural Language Processing (NLP). Sebagaimana diungkapkan oleh Ara & Rekha (2024), penerapan

teknik NLP dalam analisis lirik lagu dapat secara signifikan meningkatkan akurasi klasifikasi emosi. Untuk mencapai tujuan tersebut, penelitian ini dirancang dengan pendekatan metodologis yang terstruktur, mencakup serangkaian tahapan yang saling terkait. Tahapan-tahapan tersebut meliputi pengumpulan data (data crawling), pembersihan data (data cleaning), pra-pemrosesan teks (text pre-processing), pelabelan emosi berdasarkan leksikon, analisis N-gram, ekstraksi fitur, pelatihan model klasifikasi, dan evaluasi model. Alur keseluruhan penelitian dijelaskan secara visual pada Gambar 1.



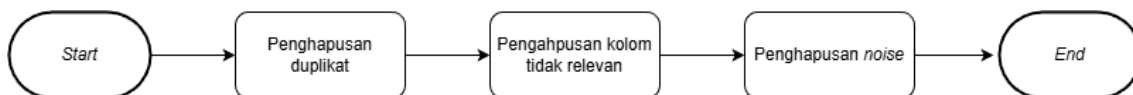
Gambar 1. Alur Proses Penelitian

2.1 Pengumpulan Data (Data Crawling)

Tahap pertama dalam penelitian ini adalah pengumpulan data. Lirik lagu diperoleh menggunakan teknik web crawling melalui Application Programming Interface (API) dari Spotify, yang merupakan metode yang telah banyak diterapkan dalam penelitian serupa (Chowdary *et al.*, 2024). Data dikumpulkan secara acak dari berbagai genre untuk memperoleh variasi emosi yang lebih luas, dengan tetap memperhatikan ketentuan *fair use* dari API Spotify. Semua data yang dikumpulkan digunakan semata-mata untuk keperluan penelitian akademik. Data yang diperoleh mencakup metadata seperti nama artis, judul lagu, serta teks lirik lengkap yang menjadi fokus analisis. Hasil dari tahap ini adalah terkumpulnya 57.650 lirik lagu berbahasa Inggris sebagai dataset awal.

2.2 Pembersihan Data (Data Cleaning)

Data yang diperoleh dari proses crawling sering kali belum terstruktur dengan baik dan mengandung elemen yang tidak relevan atau *noise*. Oleh karena itu, tahap pembersihan data sangat penting untuk memastikan kualitas dan konsistensi data sebelum dapat diproses lebih lanjut (Mishra & Chingre, 2024). Proses pembersihan ini mengikuti alur yang dapat dilihat pada Gambar 2.



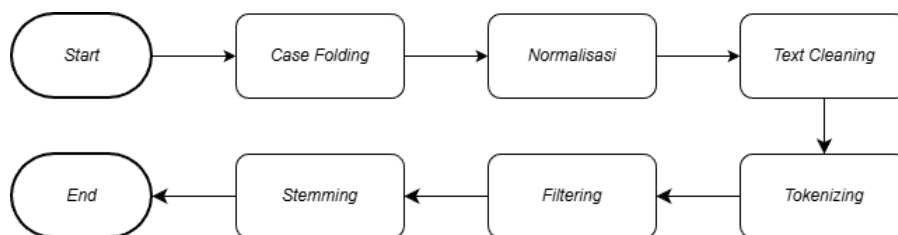
Gambar 2. Alur kerja tahapan Data Cleaning

Penghapusan duplikat untuk mengeliminasi lirik identik guna mencegah bias dalam analisis. Selanjutnya, dilakukan penghapusan kolom tidak relevan dengan menyaring metadata yang tidak diperlukan, seperti tautan eksternal. Tahap ini diakhiri dengan penghapusan *noise*, yaitu membersihkan teks lirik dari karakter atau elemen non-teks yang tidak diinginkan, seperti tag HTML atau simbol khusus, sehingga menghasilkan korpus teks yang lebih bersih dan siap untuk tahap selanjutnya.

2.3 Text Pre-processing

Setelah data dibersihkan, tahap pra-pemrosesan teks dilakukan untuk mengubah teks lirik mentah menjadi format terstruktur yang siap untuk analisis komputasi. Tahap komprehensif ini, seperti yang diilustrasikan pada Gambar 3, melibatkan serangkaian teknik NLP standar yang telah diterapkan dalam berbagai penelitian serupa (Abdillah *et al.*, 2020). Pertama, *case folding* diterapkan untuk mengonversi seluruh teks menjadi huruf kecil guna menyeragamkan data. Selanjutnya, kata-kata tidak baku dan singkatan dikonversi ke bentuk standarnya melalui normalisasi berdasarkan kamus yang telah ditetapkan. Kemudian, *text cleaning* lebih lanjut dilakukan untuk menghilangkan tanda baca,

angka, dan simbol khusus yang tidak berkontribusi dalam analisis emosi. Teks yang telah dibersihkan kemudian dipecah menjadi unit-unit kata individual atau *token* melalui proses *tokenizing*. Untuk memfokuskan analisis pada kata bermakna, kata-kata penghubung umum (*stopwords*) yang dianggap tidak memiliki muatan emosional signifikan dihilangkan dalam tahap *filtering*. Terakhir, proses *stemming* mereduksi kata-kata berimbuhan menjadi bentuk dasar atau akar katanya, sehingga dapat mengelompokkan berbagai variasi kata yang memiliki makna serupa.



Gambar 3. Alur kerja tahapan *Text Pre-processing*

2.4 Pelabelan Emosi dan Konstruksi Leksikon

Pembangunan dataset yang berlabel dengan kualitas yang baik merupakan langkah awal yang krusial dalam melatih model pembelajaran mesin berjenis *supervised*. Dalam penelitian ini, setiap lirik lagu diklasifikasikan ke dalam salah satu dari enam kategori emosi: 'happy', 'sad', 'angry', 'romantic', 'energetic', dan 'neutral'. Kategori-kategori ini diadaptasi dari taksonomi emosi dasar (seperti yang diusulkan oleh Ekman dan Plutchik) dan dikembangkan lebih lanjut untuk mencakup ekspresi yang lebih spesifik dalam lirik lagu, seperti 'romantic' (untuk mewakili kasih sayang) dan 'energetic' (untuk menggambarkan semangat dan kegembiraan). Kami menggunakan pendekatan pelabelan berbasis leksikon yang kuat dan adaptif. Leksikon emosi yang kami kembangkan merujuk pada kata-kata inti dari NRC Emotion Lexicon (Mohammad, 2023) dan WordNet-Affect (Czarnek & Stillwell, 2022), kemudian diperkaya dengan kata-kata slang dan istilah kontekstual yang sering ditemukan dalam lirik lagu (misalnya, 'danc' untuk kategori *energetic*, 'babi' untuk *romantic*). Proses validasi leksikon dilakukan dengan mengecek sampel acak untuk memastikan kesesuaian makna. Leksikon ini berisi pemetaan kata dasar (setelah dilakukan stemming) ke dalam kategori emosi yang relevan. Proses pelabelan dilakukan dengan menghitung frekuensi kemunculan kata-kata yang berasosiasi dengan setiap emosi dalam lirik. Berdasarkan perhitungan ini, setiap lirik akan diberi label pada kategori emosi yang memiliki skor tertinggi. Jika skor tertinggi adalah nol, lirik tersebut akan dikategorikan sebagai 'neutral'. Pendekatan ini belum secara eksplisit menangani negasi atau polisemi, yang merupakan keterbatasan dalam metode ini.

2.5 Analisis N-gram

Untuk mengidentifikasi pola dan frasa khas yang sering muncul dalam lirik lagu, dilakukan analisis N-gram. Metode ini, yang juga digunakan oleh Jo & Kim (2022) dalam menganalisis lirik K-Pop, berfokus pada identifikasi urutan kata yang sering muncul, mulai dari kata tunggal (unigram), pasangan kata (bigram), hingga tiga kata berurutan (trigram). Dengan menganalisis pola-pola ini untuk setiap kategori emosi, penelitian ini berusaha memperoleh pemahaman yang lebih baik tentang kombinasi kata yang secara konsisten muncul bersamaan dengan ekspresi emosional tertentu dalam lirik.

2.6 Ekstraksi Fitur

Teks yang telah diproses perlu diubah menjadi representasi numerik untuk dapat diproses oleh algoritma pembelajaran mesin. Dalam penelitian ini, kami menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk ekstraksi fitur. TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen (lirik lagu) dan sejauh mana kata tersebut jarang muncul di seluruh korpus, yang telah terbukti efektif untuk representasi teks

(Sutriawan *et al.*, 2025; Yusupov *et al.*, 2024). Dalam penelitian ini, konfigurasi TF-IDF mencakup unigram dan bigram, dengan jumlah fitur dibatasi sebanyak 1000 untuk mengoptimalkan kinerja komputasi.

2.7 Pelatihan Model Klasifikasi dan Penanganan Ketidakseimbangan Kelas

Model klasifikasi yang digunakan dalam penelitian ini adalah Random Forest Classifier. Algoritma ensemble ini dipilih karena kemampuannya yang kuat dalam menangani data dengan dimensi tinggi (seperti fitur TF-IDF), ketahanannya terhadap overfitting, serta kinerjanya yang kompetitif dalam tugas klasifikasi teks (Ngo *et al.*, 2022; Pratama *et al.*, 2025; Saptono *et al.*, 2025). Dataset yang telah diekstraksi fiturnya dibagi menjadi dua subset: data pelatihan (80%) dan data pengujian (20%) menggunakan stratified sampling untuk mempertahankan proporsi kelas yang konsisten. Model dilatih menggunakan parameter awal $n_estimators=100$, $criterion='gini'$, dan $random_state=42$. Untuk menangani masalah ketidakseimbangan kelas, penelitian ini menerapkan class weighting dengan strategi 'balanced' pada Random Forest, yang secara otomatis menyesuaikan bobot kelas berdasarkan frekuensi kemunculannya di data pelatihan.

2.8 Evaluasi Model dan Optimasi Hyperparameter

Kinerja model yang telah dilatih dievaluasi secara menyeluruh menggunakan data uji. Evaluasi dilakukan dengan menganalisis confusion matrix untuk mendapatkan gambaran visual tentang kinerja model pada setiap kelas emosi. Selanjutnya, metrik Precision, Recall, $F1-Score$, serta Macro Avg dan Weighted Avg dihitung untuk mengukur akurasi model pada setiap kelas emosi, sementara akurasi keseluruhan memberikan gambaran umum mengenai proporsi prediksi yang benar secara keseluruhan. Untuk memastikan model berfungsi secara optimal, dilakukan tuning hyperparameter menggunakan metode GridSearchCV yang dikombinasikan dengan cross-validation ($cv=5$), sebuah praktik yang juga disarankan oleh Ara & Rekha (2024). Ruang pencarian hyperparameter yang diuji meliputi: $n_estimators$ [50, 100, 200], max_depth [10, 20, None], $min_samples_split$ [2, 5, 10], dan $class_weight$ [None, 'balanced']. Proses tuning ini bertujuan untuk menemukan konfigurasi terbaik yang menghasilkan kinerja optimal berdasarkan $weighted\ avg\ F1-score$, yang selanjutnya memberikan evaluasi yang lebih andal terhadap efektivitas model dalam mengklasifikasikan emosi dalam lirik lagu.

3. Hasil dan Pembahasan

3.1 Hasil

Analisis terhadap korpus yang terdiri dari 57.494 lirik lagu menghasilkan dua temuan utama. Pertama, terkait distribusi emosi, terlihat bahwa emosi *romantic* dan *happy* mendominasi secara signifikan. Emosi *romantic* mencakup 36,2% (20.792 lagu) dari total dataset, diikuti oleh *happy* yang menyumbang 26,2% (15.072 lagu). Dominasi ini menunjukkan bahwa tema cinta, kebahagiaan, dan asmara menjadi elemen utama dalam lirik lagu yang dianalisis, sesuai dengan temuan Jo & Kim (2022) tentang kecenderungan dominasi emosi positif dalam musik populer. Emosi *sad* juga muncul dengan proporsi signifikan, mencapai 16,3% (9.398 lagu), menandakan bahwa kesedihan dan kepedihan hati tetap menjadi tema yang lazim dan diterima dalam dunia musik. Di sisi lain, emosi *energetic* yang sering diasosiasikan dengan lirik bertempo cepat dan membangkitkan semangat muncul pada 8,9% (5.124 lagu). Sebaliknya, lirik yang dikategorikan sebagai *neutral*, yang tidak mengandung kata kunci emosional jelas berdasarkan leksikon yang digunakan, hanya menyumbang 8% (4.592 lagu). Temuan yang menarik adalah bahwa emosi *angry* menempati posisi paling rendah dengan hanya 4,4% (2.516 lagu). Hal ini mengindikasikan bahwa ekspresi kemarahan, terutama dalam bentuk eksplisit, kurang populer atau mungkin tidak terlalu terepresentasi dalam genre musik yang termasuk dalam dataset ini. Kedua, dari sisi metodologi klasifikasi, model Random Forest yang digabungkan dengan teknik ekstraksi fitur TF-IDF dan penyesuaian bobot kelas menunjukkan hasil yang sangat memuaskan.

Model ini berhasil mencapai akurasi klasifikasi sebesar 83,03% setelah melalui proses optimasi hyperparameter, yang mengindikasikan efektivitas pendekatan ini dalam mengklasifikasikan emosi dalam teks lirik lagu. Proses pengumpulan data dalam penelitian ini dilakukan menggunakan Application Programming Interface (API) dari Spotify. Dengan metode ini, terkumpul sebanyak 57.650 lirik lagu. Data dikumpulkan secara acak tanpa pembatasan genre untuk mendapatkan variasi emosi yang lebih beragam. Seluruh korpus yang terkumpul menggunakan bahasa Inggris sebagai bahasa utama.

Tabel 1. Dataset Primer

id	artist	song	text	...
1	ABBA	<i>She's My Kind Of Girl</i>	<i>Look at her face, it's a wonderful face And it means something special to me Look at the way t...</i>	...
2	ABBA	<i>Andante, Andante</i>	<i>Take it easy with me, please Touch me gently like a summer evening breeze Take your time, make...</i>	...
3	ABBA	<i>As Good As New</i>	<i>I'll never know why I had to go Why I had to put up such a lousy rotten show Boy, I was tough,...</i>	...
4	ABBA	<i>Bang a boomerang</i>	<i>Making somebody happy is a question of give and take You can learn how to show it so come on, giv...</i>	...
...
57649	Zwan	<i>Desire</i>	<i>northern star am I frightened where can i go to rest i can't sleep and i'm still fightin...</i>	...
57650	Zwan	<i>Heartsong</i>	<i>come in make yourself at home i'm a bit late hate to make you wait but heart songs...</i>	...

Sebagai dasar untuk analisis, struktur dataset yang digunakan dalam penelitian ini disajikan pada Tabel 1. Tabel tersebut menggambarkan empat atribut utama yang membentuk dataset. Atribut *artist* mencatat nama pencipta atau musisi yang mempopulerkan lagu, sedangkan atribut *song* mencantumkan judul lengkap dari lagu yang dimaksud. Elemen inti dalam dataset ini terletak pada kolom *text*, yang berisi teks lirik secara lengkap, yang kemudian digunakan sebagai objek utama untuk ekstraksi dan klasifikasi emosi. Selain itu, terdapat pula kolom *link* yang mencantumkan sumber referensi dan informasi terkait hak cipta lirik lagu tersebut.

Tabel 2. Cleaned Dataset

id	artist	text
1	ABBA	<i>Look at her face, it's a wonderful face And it means something special to me Look at the way t...</i>
2	ABBA	<i>Take it easy with me, please Touch me gently like a summer evening breeze Take your time, make...</i>
3	ABBA	<i>I'll never know why I had to go Why I had to put up such a lousy rotten show Boy, I was tough,...</i>
4	ABBA	<i>Making somebody happy is a question of give and take You can learn how to show it so come on, giv...</i>
...
57493	Zwan	<i>northern star am I frightened where can i go to rest i can't sleep and i'm still fightin...</i>
57494	Zwan	<i>come in make yourself at home i'm a bit late hate to make you wait but heart songs...</i>

Tahap pra-pemrosesan data, yang divisualisasikan pada Tabel 2, berhasil menghasilkan struktur dataset yang lebih terfokus dan siap analisis. Langkah pertama yang dilakukan adalah menyederhanakan struktur data dengan hanya mempertahankan kolom *artist* dan *text*. Satu kolom atribut lain sengaja dihilangkan untuk menghindari potensi bias dalam analisis yang dapat ditimbulkan oleh metadata teknis. Selanjutnya, sebanyak 156 entri duplikat pada kolom *text* juga dihapus. Tindakan ini tidak hanya bertujuan untuk mencegah overfitting pada model, tetapi juga mengeliminasi kata kunci tidak relevan yang dapat menjadi *noise*, sehingga fokus analisis dapat diarahkan secara lebih tepat pada pola linguistik murni yang mengandung muatan emosional. Hasil akhir dari proses kurasi ini adalah sebuah dataset yang terdiri dari 57.494 data teks yang telah terstandarisasi, yang menjadi landasan yang kokoh untuk tahapan analisis emosi selanjutnya.

Tabel 3. Case Folding

id	artist	text
1	ABBA	<i>look at her face, it's a wonderful face and it means something special to me look at the way t...</i>
2	ABBA	<i>take it easy with me, please touch me gently like a summer evening breeze take your time, make...</i>
3	ABBA	<i>i'll never know why i had to go why i had to put up such a lousy rotten show boy, i was tough,...</i>
4	ABBA	<i>making somebody happy is a question of give and take you can learn how to show it so come on, giv...</i>
...		...
57493	Zwan	<i>northern star am I frightened where can i go to rest i can't sleep and i'm still fightin...</i>
57494	Zwan	<i>come in make yourself at home i'm a bit late hate to make you wait but heart songs ...</i>

Tahap *case folding* dalam penelitian ini diimplementasikan dengan memanfaatkan fungsi *lower()* dari *library Pandas* guna menormalisasi seluruh karakter dalam kolom *text* menjadi huruf kecil. Hasil dari proses ini dapat dilihat pada visualisasi Tabel 3. Pembahasan lebih lanjut mengungkap bahwa tahap ini memiliki signifikansi yang krusial. Dengan menyamakan semua variasi kapitalisasi suatu kata ke dalam bentuk huruf kecil yang seragam, tahap ini berhasil mencegah terciptanya entri-entri duplikat dalam kamus fitur. Akibatnya, kompleksitas model dapat dikurangi secara signifikan, yang pada akhirnya juga menurunkan tuntutan terhadap sumber daya komputasi yang diperlukan.

Tabel 4. Data Kamus Kata Baku

id	slang_dict	formal
1	<i>ain't</i>	<i>am not</i>
2	<i>aren't</i>	<i>are not</i>
3	<i>can't</i>	<i>cannot</i>
4	<i>couldn't</i>	<i>could not</i>
...
86541	<i>gotta</i>	<i>got to</i>
86542	<i>kinda</i>	<i>kind of</i>

Untuk menangani karakteristik bahasa yang tidak formal dalam lirik lagu, dilakukan proses normalisasi dengan memanfaatkan kamus kata slang (*slang_dict*). Kamus ini berfungsi sebagai acuan untuk mentransformasikan istilah-istilah informal misalnya "*kinda*" menjadi bentuk standarnya "*kind of*", sebagaimana disajikan pada Tabel 3. Mekanisme penerapannya diawali dengan tokenisasi teks menjadi per kata. Setiap token kemudian dicocokkan dengan kunci dalam *slang_dict*. Apabila

kecocokan ditemukan, kata tersebut diganti dengan nilai bakunya. Akhirnya, teks yang telah dibakukan dibentuk kembali dari kumpulan kata yang telah diproses. Normalisasi ini merupakan fondasi penting yang meningkatkan reliabilitas fitur linguistik untuk pelatihan model.

Tabel 5. Hasil Setelah Normalisasi

id	artist	text
1	ABBA	<i>look at her face, it is a wonderful face and it means something special to me look at the way t...</i>
2	ABBA	<i>take it easy with me, please touch me gently like a summer evening breeze take your time, make...</i>
3	ABBA	<i>i will never know why i had to go why i had to put up such a lousy rotten show boy, i was tough,...</i>
4	ABBA	<i>making somebody happy is a question of give and take you can learn how to show it so come on, giv...</i>
...
57493	Zwan	<i>northern star am I frightened where can i go to rest i cannot sleep and i am still fightin...</i>
57494	Zwan	<i>come in make yourself at home i am a bit late hate to make you wait but heart songs...</i>

Hasil proses normalisasi kata slang disajikan secara visual dalam Tabel 5. Implementasi tahap ini terbukti efektif untuk menyeragamkan variasi penulisan nonbaku, seperti singkatan, yang lazim dijumpai dalam korpus lirik lagu. Tahap normalisasi memainkan peran krusial sebagai prasyarat untuk meminimalisir noise sebelum analisis lebih lanjut, sehingga secara signifikan meningkatkan akurasi dalam mendeteksi pola linguistik yang mengandung muatan emosi. Hal ini penting mengingat konsistensi kosakata merupakan faktor penentu bagi validitas interpretasi fitur teks yang dihasilkan.

Tabel 6. Hasil Cleaning Test

id	artist	text
1	ABBA	<i>look at her face it is a wonderful face and it means something special to me look at the way t...</i>
2	ABBA	<i>take it easy with me please touch me gently like a summer evening breeze take your time make...</i>
3	ABBA	<i>i will never know why i had to go why i had to put up such a lousy rotten show boy i was tough...</i>
4	ABBA	<i>making somebody happy is a question of give and take you can learn how to show it so come on giv...</i>
...
57493	Zwan	<i>northern star am I frightened where can i go to rest i cannot sleep and i am still fightin...</i>
57494	Zwan	<i>come in make yourself at home i am a bit late hate to make you wait but heart songs...</i>

Hasil proses *text cleaning* yang divisualisasikan pada Tabel 6, diawali dengan tahap pembersihan teks untuk memurnikan lirik dari elemen non-linguistik seperti angka, tanda baca, dan simbol. Pemurnian ini dilakukan menggunakan ekspresi reguler (*regex*), di mana pola $r'^{[\wedge a-zA-Z \s]}$ berfungsi menghapus semua karakter yang bukan huruf atau spasi, sementara pola $r'^{\s+}$ dan metode *.strip()* digunakan untuk menormalisasi spasi berlebih serta membersihkan spasi di ujung string, sehingga menghasilkan teks yang seragam dan konsisten. Tahap fundamental ini sangat krusial karena dengan mengeliminasi "noise" yang tidak mengandung muatan emosional, analisis NLP dapat

lebih terfokus pada konten linguistik murni yang merepresentasikan perasaan, sekaligus memastikan konsistensi data yang memungkinkan algoritma mengidentifikasi dan menginterpretasikan setiap kata dengan lebih akurat.

Tabel 7. Hasil Tokenisasi Kata

id	artist	text
1	ABBA	['look', 'at', 'her', 'face', 'it', 'is', 'a', 'wonderful', 'face', 'and', 'it', 'means', 'something', 'special', 'to', 'me', 'look', 'at', 'the'...
2	ABBA	['take', 'it', 'easy', 'with', 'me', 'please', 'touch', 'me', 'gently', 'like', 'a', 'summer', 'evening', 'breeze', 'take', 'your', 'time', 'make'...
3	ABBA	['i', 'will', 'never', 'know', 'why', 'i', 'bad', 'to', 'go', 'why', 'i', 'bad', 'to', 'put', 'up', 'such', 'a', 'lousy', 'rotten', 'show'...
4	ABBA	['making', 'somebody', 'happy', 'is', 'a', 'question', 'of', 'give', 'and', 'take', 'you', 'can', 'learn', 'how', 'to', 'show', 'it', 'so', 'come', 'on'...
...
57493	Zwan	['northern', 'star', 'am', 'i', 'frightened', 'where', 'can', 'i', 'go', 'to', 'rest', 'i', 'can', 'not', 'sleep', 'and', 'i', 'am', 'still', 'fighting'...
57494	Zwan	['come', 'in', 'make', 'yourself', 'at', 'home', 'i', 'am', 'a', 'bit', 'late', 'hate', 'to', 'make', 'you', 'wait', 'but', 'heart', 'songs', 'are'...

Tahap tokenisasi dalam penelitian ini dilakukan dengan memanfaatkan fungsi *word_tokenize* dari *library* NLTK (*Natural Language Toolkit*). Fungsi ini diaplikasikan pada seluruh korpus teks yang telah melalui tahap pembersihan (*text cleaning*). Untuk memastikan konsistensi proses, digunakan *Punkt Tokenizer* sebagai model referensi, dan sebuah fungsi pembungkus (*wrapper*), yakni *word_tokenize_wrapper*, dibuat untuk mengaplikasikan proses tokenisasi secara efisien ke seluruh dataset menggunakan metode *apply()* pada *Pandas*. Hasil dari proses ini mengonversi setiap dokumen teks (lirik lagu) menjadi sekumpulan *token* individual dalam bentuk *list*, yang selanjutnya dapat diamati pada berikut.

Tabel 8. Hasil Filtering Kata

id	artist	text
1	ABBA	['look', 'face', 'wonderful', 'face', 'means', 'something', 'special', 'look'...
2	ABBA	['take', 'easy', 'please', 'touch', 'gently', 'like', 'summer', 'evening', 'breeze', 'take', 'time', 'make'...
3	ABBA	['never', 'know', 'go', 'put', 'lousy', 'rotten', 'show'...
4	ABBA	['making', 'somebody', 'happy', 'question', 'give', 'take', 'learn', 'show', 'come' ...
...
57493	Zwan	['northern', 'star', 'frightened', 'go', 'rest', 'sleep', 'still', 'fighting'...
57494	Zwan	['come', 'make', 'home', 'bit', 'late', 'hate', 'make', 'wait', 'heart', 'songs'...

Sebagaimana divisualisasikan dalam Tabel 8, tahap *filtering* berhasil menyaring *stop words* bahasa inggris dari korpus teks yang dimuat dari *nlTK.corpus.stopwords* dalam *library* NLTK, sehingga menghasilkan sejumlah *token* kunci yang lebih mampu mencerminkan nuansa emosional. Proses eliminasi partikel non-esensial ini, yang dioptimasi dengan mengonversi daftar *stop words* ke dalam bentuk set, terbukti mereduksi redundansi kata secara signifikan. Berkat penerapan fungsi *stop_removal*, korpus teks yang dihasilkan telah terbebas dari *noise* linguistik, sehingga memungkinkan algoritma untuk lebih terfokus pada leksikon diagnostik penanda emosi. Dengan demikian, tahap ini tidak hanya menyederhanakan data, tetapi juga berkontribusi langsung terhadap peningkatan akurasi identifikasi emosi dalam analisis lebih lanjut.

Tabel 9. Hasil Stemming Kata

id	artist	text
1	ABBA	['look', 'face', 'wonder', 'face', 'mean', 'someth', 'special', 'look'...
2	ABBA	['take', 'easi', 'pleas', 'touch', 'gentli', 'like', 'summer', 'even', 'breez', 'take', 'time', 'make'...
3	ABBA	['never', 'know', 'go', 'put', 'lousi', 'rotten', 'show'...
4	ABBA	['make', 'somebodi', 'happi', 'question', 'give', 'take', 'learn', 'show', 'come'...
...
57493	Zwan	['northern', 'star', 'frighten', 'go', 'rest', 'sleep', 'still', 'fight'...
57494	Zwan	['come', 'make', 'home', 'bit' late', 'bate', 'make', 'wait', 'heart', 'song'...

Pada penelitian ini, proses *stemming* diimplementasikan dengan menggunakan algoritma *PorterStemmer* untuk mereduksi kata-kata berimbuhan menjadi bentuk dasar atau akar katanya. Implementasi ini memberikan dampak ganda yang signifikan, baik dari aspek komputasional maupun linguistik. Secara komputasional, teknik ini berhasil menyederhanakan model dengan mengurangi dimensionalitas ruang fitur, yang berimplikasi pada memangkas waktu pelatihan model dan menekan potensi *overfitting*. Dari perspektif semantik yang lebih krusial dalam analisis emosi, *stemming* berfungsi untuk mengonsolidasikan makna kata. Berbagai variasi morfologis dari sebuah konsep (misalnya, 'happi', 'happy', 'happiness') dikelompokkan ke dalam satu leksem dasar. Konsolidasi ini memastikan bahwa intensitas emosional dari suatu konsep tidak terfragmentasi ke dalam beberapa representasi yang berbeda, sehingga model menjadi lebih konsisten dan akurat dalam mendeteksi muatan afektif dalam lirik. Efektivitas dari proses *stemming* ini dapat dilihat pada Tabel 9.

Kode Program 1. Kamus Emosi

```
emotion_lexicon = {
    'happy': ['love', 'happi', 'joy', 'smile', 'laugh', 'fun', 'beauti',
'wonder', 'amaz'],
    'sad': ['sad', 'cri', 'tear', 'loneli', 'miss', 'hurt', 'pain', 'blue',
'broken'],
    'angry': ['angri', 'mad', 'hate', 'rage', 'furious', 'annoy', 'frustrat'],
    'romantic': ['love', 'heart', 'kiss', 'touch', 'hold', 'darling', 'babi',
'sweet'],
    'energetic': ['danc', 'move', 'rock', 'parti', 'energi', 'fire', 'burn',
'wild']
}
```

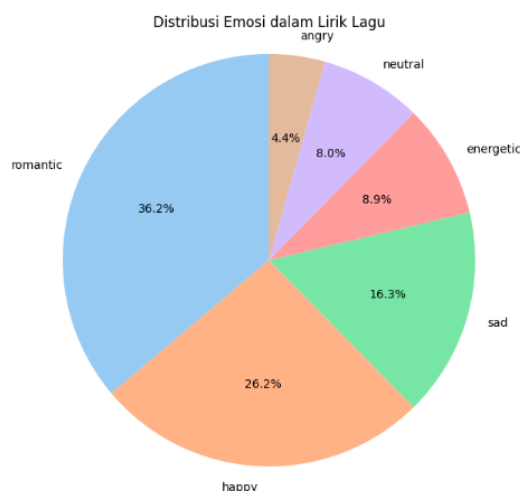
Selanjutnya adalah pelabelan emosi, yang bertujuan untuk mengkategorikan setiap lirik lagu ke dalam salah satu dari enam kelas emosi yang telah ditetapkan, yakni 'happy', 'sad', 'angry', 'romantic', 'energetic', dan 'neutral'. Untuk mencapai tujuan ini, penelitian ini mengadopsi pendekatan berbasis kamus emosi (*emotion lexicon*). Dalam praktiknya, sebuah kamus emosi (*emotion_lexicon*) dikompilasi, yang berisi daftar kata kunci dalam bentuk *stem* atau akar kata untuk merepresentasikan setiap kategori emosi. *Lexicon* ini kemudian berfungsi sebagai landasan referensi utama dalam proses identifikasi dan anotasi nuansa emosional yang terkandung dalam korpus teks lirik, sebagaimana terlihat pada Kode Program 1 yang menyertai penelitian ini.

Tabel 10. Hasil Keluaran Proses Labelling

id	emotion_label	jumlah
1	romantic	20792
2	happy	15072
3	sad	9398

4	<i>energetic</i>	5124
5	<i>neutral</i>	4592
6	<i>angry</i>	2516

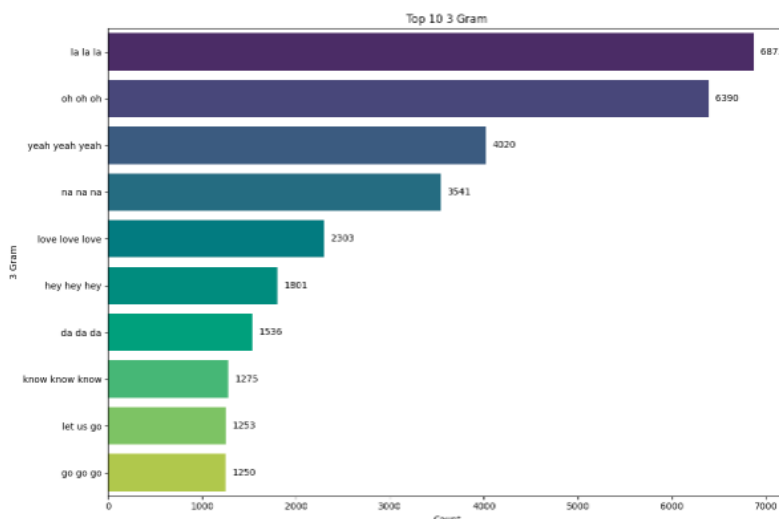
Tahap pemberian label emosi pada lirik yang telah melalui pra-pemrosesan dilakukan dengan menerapkan suatu prosedur komputasi yang sistematis. Algoritma yang dirancang memulai prosesnya dengan menginisialisasi semua skor emosi pada nilai 0. Setiap kata dalam lirik kemudian ditelusuri dan dikonfrontasikan dengan sebuah leksikon emosi dalam *emotion_lexicon* (Kode Program 1). Sebagai ilustrasi, ketika sebuah kata seperti *'happi'* teridentifikasi sesuai dengan entri dalam leksikon untuk emosi *'happy'*, maka skor untuk kategori emosi tersebut akan dinaikkan sebesar 1 poin. Setelah proses penelusuran seluruh kata selesai, algoritma melakukan perbandingan terhadap skor akhir dari setiap kategori emosi untuk mengidentifikasi nilai tertinggi (*max_score*). Label emosi dominan kemudian ditetapkan berdasarkan perbandingan ini. Dalam situasi di mana *max_score* bernilai nol ($=0$) yang mengindikasikan tidak adanya kata kunci emosional yang terdeteksi lirik tersebut secara otomatis dikategorikan ke dalam label *'netral'*. Sebaliknya, jika *max_score* lebih besar dari nol (>0), label yang ditetapkan adalah emosi pertama yang mencapai skor tertinggi tersebut. Sebagai output akhir dari prosedur ini, label emosi untuk setiap lirik kemudian diintegrasikan ke dalam kerangka data utama dengan menambah kolom baru bernama *emotion_label* pada *DataFrame df_clean*. Distribusi frekuensi dari setiap kategori emosi yang dihasilkan dari proses ini dapat diamati secara lebih rinci pada Tabel 10.



Gambar 4. Hasil Visualisasi Distribusi Emosi Lirik Lagu dengan Pie Chart

Hasil analisis distribusi sentimen terhadap 57.490 lirik lagu disajikan dalam bentuk Pie Chart pada Gambar 4. Visualisasi ini dibangun berdasarkan analisis statistik deskriptif yang menunjukkan komposisi proporsi setiap kategori emosi yang teridentifikasi. Kuantifikasi dilakukan dengan menghitung frekuensi kemunculan setiap emosi menggunakan metode *value_counts()*. Berdasarkan hasil tersebut, teramati dominasi signifikan dari dua emosi positif, yaitu *romantic* dan *happy*. Emosi *romantic* mencakup 36,2% (20.792 lagu) dari total dataset, diikuti oleh *happy* yang menyumbang 26,2% (15.072 lagu). Temuan ini mencerminkan bahwa tema cinta, kebahagiaan, dan asmara menjadi inti dari lirik lagu yang dianalisis, sejalan dengan hasil penelitian Jo & Kim (2022) tentang dominasi emosi positif dalam musik populer. Di sisi lain, emosi *sad* juga menempati proporsi substansial, yakni 16,3% (9.398 lagu), yang menunjukkan bahwa kesedihan tetap menjadi tema sentral dalam musik, berfungsi sebagai sarana ekspresi emosional yang dapat menghubungkan pendengar. Emosi *energetic*, yang umumnya dikaitkan dengan lirik bertempo cepat dan membangkitkan semangat, terdistribusi pada 8,9% (5.124 lagu).

Lirik yang dikategorikan *neutral*, yaitu yang tidak memuat kata kunci emosional berdasarkan leksikon yang digunakan, hanya mencakup 8% (4.592 lagu). Menariknya, emosi *angry* menempati kategori dengan persentase terendah, yakni hanya 4,4% (2.516 lagu), yang dapat mengindikasikan bahwa ekspresi kemarahan lebih jarang ditemukan atau kurang diterima dalam lirik musik populer, atau karena keterbatasan leksikon dalam mengidentifikasi nuansa kemarahan yang lebih halus. Untuk menelaah makna kontekstual dan pola linguistik dalam lirik lagu, penelitian ini menerapkan pendekatan analisis N-gram. Metode ini memungkinkan eksplorasi struktur leksikal pada berbagai tingkat granularitas, dari kata tunggal (unigram) hingga frasa yang lebih kompleks seperti bigram dan trigram. Pada tingkat pertama, analisis unigram mengidentifikasi kata-kata dengan frekuensi tertinggi, seperti "love", "go", dan "know", yang memberikan gambaran umum mengenai leksikon emosional yang dominan dalam korpus. Namun, analisis unigram memiliki keterbatasan karena tidak memperhitungkan hubungan antar kata. Oleh karena itu, analisis berlanjut pada bigram dan trigram untuk menangkap nuansa yang lebih dalam. Analisis bigram mengidentifikasi pasangan kata berurutan (misalnya, "oh oh", "yeah yeah") yang sering kali memperkuat nuansa perasaan atau memberikan kesan mantrik dalam lirik. Selanjutnya, eksplorasi trigram (urutan tiga kata) seperti pola "la la la" mengungkapkan konteks yang lebih kaya, tidak hanya memperkuat ekspresi emosional tetapi juga mencerminkan struktur musikal dan gaya khas dalam lirik lagu. Kombinasi ketiga tingkat analisis ini memungkinkan pemahaman yang lebih luas tentang perkembangan ekspresi emosi, dari level kata hingga kolokasi kata yang membentuk ekspresi lebih kompleks. Dalam penelitian ini, analisis N-gram diimplementasikan menggunakan *Natural Language Toolkit* (NLTK) untuk tahap pra-pemrosesan teks, yang meliputi penyaringan stopwords dalam bahasa Inggris untuk memfokuskan analisis pada kata-kata yang lebih substansial. Korpus lirik kemudian diubah menjadi representasi numerik menggunakan *CountVectorizer* dari library *scikit-learn*. Teknik ini menghitung frekuensi kemunculan setiap trigram, yang kemudian diurutkan untuk mengidentifikasi pola-pola yang paling signifikan. Hasil implementasi ini berhasil mengidentifikasi sepuluh trigram dengan frekuensi tertinggi, yang divisualisasikan dalam Gambar 5. Pola-pola leksikal yang teridentifikasi ini menjadi dasar empiris untuk membahas hubungan antara struktur linguistik dan muatan emosional dalam lirik lagu pada bagian pembahasan berikutnya.



Gambar 5. Hasil Visualisasi Trigram Lirik Lagu

Proses pemisahan dataset dilakukan dengan menerapkan *stratified random sampling* guna menjawab tantangan ketidakseimbangan distribusi emosi dalam korpus lirik. Rasio 80:20 dialokasikan untuk data latih dan data uji, yang tidak hanya merupakan pembagian kuantitatif, tetapi juga sebuah strategi untuk mempertahankan proporsi setiap kelas emosi pada kedua subset. Hasil dari implementasi teknik ini

berhasil menghasilkan 45.995 sampel untuk pelatihan dan 11.499 sampel untuk pengujian. Yang patut diperhatikan, pendekatan *stratified* ini terbukti krusial untuk meminimalisasi bias selama pelatihan model, sehingga model tidak didominasi oleh emosi dengan frekuensi tertinggi. Lebih lanjut, untuk memastikan bahwa temuan dalam studi ini dapat direplikasi, parameter `random_state` diatur pada nilai 42. Konfigurasi ini berfungsi sebagai "kunci" yang mengunci proses pengacakan, menjamin konsistensi pembagian data yang sama pada setiap eksekusi, yang merupakan fondasi bagi reproduktibilitas eksperimen.

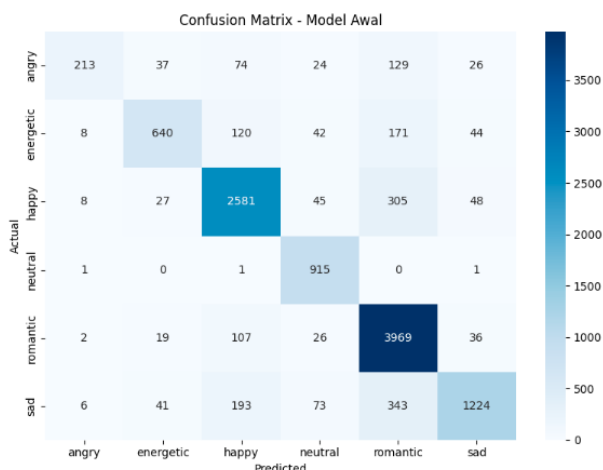
Kode Program 2. *Feature Extraction* dengan TF-IDF

```
vectorizer = TfidfVectorizer(max_features=1000, ngram_range=(1, 2))  
X = vectorizer.fit_transform(df_clean['processed_text'])  
y = df_clean['emotion_label']  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,  
random_state=42, stratify=y)  
classifier = RandomForestClassifier(n_estimators=100, random_state=42)  
classifier.fit(X_train, y_train)
```

Pada tahap inti pemrosesan, penelitian ini menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengekstraksi fitur dari data teks. Proses ini, yang diterapkan dalam Kode Program 2, bertujuan mengubah lirik lagu yang telah dibersihkan (tersimpan dalam kolom *processed_text*) dari bentuk teks mentah menjadi representasi numerik yang dapat diproses oleh algoritma pembelajaran mesin. Untuk mengoptimalkan representasi fitur, *TF-IDF Vectorizer* dikonfigurasi dengan membatasi jumlah fitur menjadi 1.000 term dengan bobot tertinggi. Pendekatan ini memungkinkan model untuk menangkap makna kata secara individu, sekaligus mempertimbangkan konteks lokal dalam frasa-frasa tertentu dalam lirik. Selanjutnya, untuk mengevaluasi kinerja model secara objektif, dataset dibagi menjadi dua subset: data pelatihan (80%) dan data pengujian (20%), dengan penerapan *stratified sampling* untuk memastikan distribusi label emosi yang proporsional pada kedua subset tersebut, agar komposisi data tetap merepresentasikan data asli. Pembagian ini bertujuan untuk menghindari bias dalam proses pelatihan dan pengujian. Pada tahap berikutnya, model *Random Forest Classifier* dibangun untuk melakukan klasifikasi emosi. Dengan menggunakan library *sklearn.ensemble* dan parameter `n_estimators=100`, `criterion='entropy'`, `random_state=42`, dan `class_weight='balanced'`, model dilatih menggunakan data pelatihan yang terdiri dari vektor *TF-IDF* (X_{train}) dan label sentimen (y_{train}).

Evaluasi awal terhadap model *Natural Language Processing* (NLP) yang dikembangkan menunjukkan tingkat akurasi keseluruhan sebesar 82,98%. Hasil ini menunjukkan bahwa model memiliki kinerja yang solid, meskipun analisis mendalam terhadap metrik klasifikasi per kelas mengungkapkan variasi performa yang signifikan antara kategori emosi. Model menunjukkan kemampuan yang sangat baik dalam mengidentifikasi lagu dengan emosi *neutral*, dengan nilai *recall* sempurna sebesar 1,00. Ini berarti seluruh lagu yang benar-benar bertema *neutral* berhasil dikenali oleh model. Namun, nilai *precision* yang hanya mencapai 0,81 menunjukkan bahwa sekitar 19% lagu yang diklasifikasikan sebagai *neutral* ternyata merupakan emosi lain, mengindikasikan adanya sedikit kontaminasi pada klasifikasi kategori ini. Pada kelas *romantic*, model menunjukkan kemampuan impresif dengan *recall* 0,95, yang berarti hampir semua lagu *romantic* terdeteksi. Namun, dengan nilai *precision* yang sama dengan kelas *neutral* (0,81), model cenderung melakukan over-klasifikasi atau terlalu longgar dalam memberi label *romantic*, sehingga beberapa lagu dari kategori emosi lain terkategori sebagai *romantic*. Di sisi lain, model mengalami kesulitan dalam mendeteksi emosi *angry*. Nilai *recall* yang rendah (0,42) dan *F1-score* yang hanya 0,57 menunjukkan bahwa model gagal mengenali sebagian besar lagu yang seharusnya bernuansa kemarahan. Hal ini menyoroti bahwa lebih dari separuh lagu yang sebenarnya mengandung kemarahan tidak terdeteksi dengan benar. Sementara itu, performa model pada kelas *energetic* dan *sad* berada pada tingkat menengah, dengan *recall* masing-masing 0,62 dan 0,65, menunjukkan adanya ruang untuk peningkatan. Setelah menganalisis pola kesalahan klasifikasi ini, penyebab utama yang diduga adalah ketidakseimbangan distribusi data dalam dataset. Kelas mayoritas seperti *romantic*

diwakili oleh 4.159 sampel, sedangkan kelas minoritas seperti *angry* hanya terdiri dari 503 sampel. Ketidakseimbangan seperti ini sering kali menyebabkan model menjadi bias, karena algoritma lebih mudah mempelajari pola-pola dari kelas dengan representasi data yang lebih banyak, sementara mengorbankan akurasi pada kelas dengan sampel yang lebih sedikit. Inilah yang diperkirakan menjadi alasan utama mengapa model kesulitan mengenali emosi *angry* dan lebih cenderung mengklasifikasikan lagu ke dalam kategori *romantic*.



Gambar 6. Visualisasi *Confusion Matrix*

Visualisasi pada Gambar 6 menggambarkan *confusion matrix* yang digunakan untuk menganalisis kinerja model klasifikasi emosi dalam lirik lagu. Matriks ini, yang dibuat dengan memanfaatkan *sklearn.metrics.confusion_matrix*, memberikan gambaran yang jelas dengan mengkategorikan hasil prediksi ke dalam empat kuadran: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Dari kategori-kategori dasar ini, metrik evaluasi seperti akurasi, presisi, recall, dan *F1-score* dihitung, memberikan tolok ukur kuantitatif terhadap efektivitas model dalam menangkap pola linguistik yang mewakili emosi. Pembahasan lebih lanjut mengenai *confusion matrix* ini difokuskan untuk mengidentifikasi kekuatan dan kelemahan model, terutama terkait tantangan *class imbalance*. Model menunjukkan performa yang sangat baik pada kelas *neutral* (dengan *recall* 1,00) dan *romantic* (dengan *recall* 0,95), namun kesulitan dalam mendeteksi emosi *angry*, dengan *recall* hanya 0,42. Hal ini terlihat jelas pada banyaknya instance *angry* yang sebenarnya (FN=290) salah diklasifikasikan sebagai *romantic* (TP=213) atau kategori lain. Pola kesalahan ini mengindikasikan bahwa fitur leksikal untuk *angry* mungkin tumpang tindih dengan kata-kata intensitas tinggi yang juga sering muncul dalam konteks *romantic* atau *energetic*. Selain itu, bisa jadi leksikon yang digunakan kurang mencakup variasi ekspresi kemarahan dalam lirik lagu.

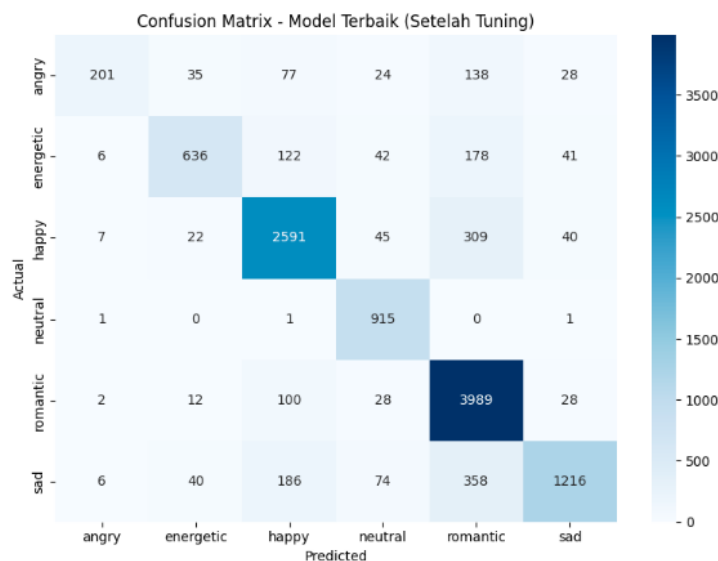
Tabel 11. *Classification Report*

	Precision	Recall	F1-Score	Support
angry	0.89	0.42	0.57	503
energetic	0.84	0.62	0.72	1025
happy	0.84	0.86	0.85	3014
neutral	0.81	1.00	0.90	918
romantic	0.81	0.95	0.87	4159
sad	0.89	0.65	0.75	1880
accuracy			0.83	11499
macro avg	0.85	0.75	0.78	11499
weighted avg	0.84	0.83	0.82	11499

Hasil evaluasi model yang disajikan dalam Tabel 11 menunjukkan bahwa model secara keseluruhan berhasil mencapai kinerja yang solid, dengan akurasi sebesar 82,98% dan *weighted avg F1-score* 0,82, yang mengindikasikan kemampuannya dalam mengidentifikasi emosi dalam lirik lagu. Namun, analisis lebih mendalam mengungkapkan adanya variasi performa antara kelas emosi yang berbeda. Model menunjukkan hasil yang sangat baik dalam mengidentifikasi emosi *happy*, *romantic*, dan *neutral*, dengan nilai presisi dan recall yang seimbang serta tinggi. Sebaliknya, untuk emosi *angry*, model menunjukkan ketidakseimbangan metrik: meskipun presisinya tinggi (0,89), *recall*-nya sangat rendah (0,42). Hal ini berarti, meskipun prediksi *angry* yang dilakukan oleh model hampir selalu benar, model sering kali melewatkan lirik yang seharusnya dikategorikan sebagai *angry*. Rendahnya *macro avg F1-score* (0,78) dibandingkan dengan *weighted avg F1-score* (0,82) mengonfirmasi bahwa performa model pada kelas minoritas seperti *angry* dan *energetic* masih perlu ditingkatkan. Dalam analisis kesalahan (*error analysis*), pola kesalahan klasifikasi yang ditemukan menunjukkan beberapa kekurangan. Sebagai contoh, lirik "fire burning in my soul, can't control this rage no more" diklasifikasikan sebagai *energetic* (karena kata *fire* dan *burn* terasosiasi dengan leksikon *energetic*), padahal dalam konteks ini, lirik tersebut lebih tepat dikategorikan sebagai *angry*. Contoh lainnya, lirik "I hate the way you lie, I hate the way you cry" diklasifikasikan sebagai *sad* (karena kata *cry*), sementara kata *hate* seharusnya lebih kuat menunjukkan emosi *angry*. Hal ini menunjukkan kelemahan dari pendekatan leksikon berbasis frekuensi kata, yang tidak cukup memperhatikan konteks kalimat, negasi, atau penggunaan intensifier.

Tabel 12. *Tuned Model Classification Report*

	Precision	Recall	F1-Score	Support
angry	0.90	0.40	0.55	503
energetic	0.85	0.62	0.72	1025
happy	0.84	0.86	0.85	3014
neutral	0.81	1.00	0.89	918
romantic	0.80	0.96	0.87	4159
sad	0.90	0.65	0.75	1880
accuracy			0.83	11499
macro avg	0.85	0.75	0.77	11499
weighted avg	0.84	0.83	0.82	11499



Gambar 7. Visualisasi *Confusion Matrix* Setelah *Tuning*

Hasil optimasi hyperparameter yang dilakukan melalui metode *GridSearchCV* berhasil menentukan konfigurasi optimal untuk model: $n_estimators=200$, $max_depth=20$, $min_samples_split=5$, dan $class_weight='balanced'$. Hasil konfigurasi ini dapat dilihat pada Tabel 12 dan Gambar 7. Evaluasi menggunakan *cross-validation* ($cv=5$) menghasilkan skor terbaik sebesar 82,59%. Setelah model dikonfigurasi dengan parameter terbaik ini dan diuji pada dataset uji, akurasi akhir yang diperoleh adalah 83,03%. Peningkatan ini sangat marginal, hanya 0,03% dibandingkan dengan kinerja model baseline sebelum tuning (82,98%). Peningkatan yang minimal ini mengindikasikan bahwa model baseline dengan *class weighting* telah beroperasi pada level kinerja yang hampir optimal untuk konfigurasi algoritma dan fitur yang digunakan. Dengan kata lain, ruang untuk peningkatan lebih lanjut melalui optimasi hyperparameter pada pipeline ini tampaknya sangat terbatas. Temuan ini memperkuat argumen bahwa masalah utama terletak pada representasi data (ketidakseimbangan kelas dan keterbatasan fitur leksikal), bukan pada penyediaan parameter model *Random Forest*. Pembahasan lebih lanjut menghubungkan temuan ini dengan studi lain. Dominasi emosi positif (*romantic* dan *happy*) sejalan dengan temuan Jo & Kim (2022) dalam lirik K-Pop, yang mungkin mencerminkan preferensi audiens global atau fungsi musik sebagai sarana pelipur lara. Rendahnya representasi emosi *angry* dapat disebabkan oleh ekspresi kemarahan yang lebih banyak disalurkan melalui elemen musikal (seperti distorsi gitar atau tempo cepat) ketimbang lirik secara eksplisit, atau karena adanya stigma sosial terhadap kemarahan. Beberapa keterbatasan dalam penelitian ini antara lain: (1) Generalisasi terbatas pada lirik berbahasa Inggris dari Spotify, (2) Tidak adanya validasi manual terhadap label yang dihasilkan oleh leksikon, (3) Leksikon kustom yang digunakan belum divalidasi secara eksternal dan rentan terhadap subjektivitas, (4) Belum menangani negasi, sarkasme, atau metafora kompleks, (5) Tidak membandingkan performa model dengan model baseline lain (seperti *SVM*, *Naive Bayes*) atau model deep learning untuk klaim "kompetitif". Untuk aplikasi praktis seperti sistem rekomendasi, model dengan *recall* rendah untuk *angry* dapat menyebabkan kurangnya keragaman dalam rekomendasi bagi pengguna yang menyukai musik dengan nuansa kemarahan.

3.2 Pembahasan

Hasil penelitian ini menunjukkan bahwa emosi positif, khususnya 'romantic' dan 'happy', mendominasi lirik lagu populer berbahasa Inggris yang dianalisis. Temuan ini sejalan dengan hasil penelitian Jo & Kim (2022), yang menemukan bahwa emosi positif seperti kebahagiaan dan percintaan sering kali menjadi tema utama dalam lirik lagu K-Pop, mencerminkan kecenderungan pasar dan preferensi audiens global terhadap tema-tema yang mengangkat perasaan positif. Dominasi emosi positif ini juga konsisten dengan temuan yang dilakukan oleh Ara & Rekha (2024) dalam analisis lirik lagu yang menunjukkan bahwa perasaan cinta dan kebahagiaan sering menjadi narasi sentral dalam berbagai genre musik, seiring dengan dampaknya terhadap pengalaman emosional pendengar. Namun, meskipun tema-tema positif mendominasi, emosi 'sad' atau kesedihan tetap memiliki porsi signifikan, yakni 16,3%. Hal ini memperlihatkan bahwa ekspresi kesedihan tetap penting dalam ekspresi musik, berfungsi sebagai media katarsis bagi pendengar yang mengalami atau mengidentifikasi dengan perasaan tersebut. Temuan ini selaras dengan hasil yang ditemukan oleh Bužić & Dobša (2025) dalam studi mengenai lirik Eurovision, yang juga menunjukkan keberadaan emosi negatif seperti kesedihan dan rasa kehilangan, meskipun tidak dominan. Dalam konteks ini, kesedihan dalam lirik lagu dapat berfungsi sebagai penghubung emosional dengan pendengar, menjadikannya bagian integral dari narasi dalam musik populer. Penelitian ini juga menemukan bahwa emosi 'angry' atau kemarahan merupakan kategori yang sangat jarang ditemukan, hanya mencakup 4,4% dari keseluruhan dataset. Temuan ini mengindikasikan adanya ketidakseimbangan kelas yang signifikan, di mana emosi yang lebih jarang muncul dalam lirik lagu populer menjadi lebih sulit untuk teridentifikasi. Hal ini sejalan dengan penelitian yang dilakukan oleh Jo & Kim (2022), yang juga mencatat bahwa kemarahan dalam lirik musik, terutama di genre populer, tidak terlalu sering dieksplorasi dibandingkan dengan emosi positif. Kemungkinan, hal ini disebabkan oleh kecenderungan pasar yang lebih menerima tema-tema emosional yang lebih ringan dan menyenangkan.

Selain itu, keterbatasan leksikon dalam menangkap nuansa emosional seperti kemarahan, terutama yang tidak langsung diekspresikan dalam kata-kata eksplisit, dapat memperburuk masalah ini. Dari sisi metodologi, pendekatan menggunakan model *Random Forest* dengan teknik ekstraksi fitur *TF-IDF* dan leksikon emosi kustom terbukti efektif dalam mengklasifikasikan emosi dalam lirik lagu dengan akurasi 83,03%. Model ini menunjukkan hasil yang cukup baik, namun masih menghadapi tantangan dalam mendeteksi emosi minoritas, khususnya 'angry' dan 'energetic'. Hal ini menunjukkan bahwa ketidakseimbangan distribusi data menjadi isu utama yang mempengaruhi kinerja model. Temuan ini memperkuat argumen yang diajukan oleh Ara & Rekha (2024) bahwa pendekatan multi-fitur NLP dan pemodelan berbasis ensemble seperti *Random Forest* bisa menghasilkan model yang solid, meskipun tetap memerlukan penyesuaian dalam menangani masalah ketidakseimbangan kelas. Untuk itu, penelitian lebih lanjut disarankan untuk mengeksplorasi teknik penyeimbangan kelas yang lebih canggih, seperti *SMOTE* atau *cost-sensitive learning*, untuk mengatasi masalah ini. Secara keseluruhan, hasil penelitian ini menyarankan bahwa meskipun model berbasis *Random Forest* dengan *TF-IDF* dan leksikon emosi kustom cukup efektif, masih terdapat ruang untuk perbaikan, terutama dalam menangani ketidakseimbangan kelas dan memperkaya leksikon emosi agar dapat menangkap konteks lebih luas dari lirik lagu. Selain itu, penggunaan model berbasis *deep learning* seperti *BERT*, yang dapat menangkap hubungan kontekstual antar kata dengan lebih baik, berpotensi meningkatkan hasil klasifikasi emosi dalam lirik lagu di masa depan.

4. Kesimpulan

Berdasarkan analisis menyeluruh terhadap korpus yang terdiri dari 57.494 lirik lagu berbahasa Inggris, penelitian ini berhasil mengungkap dua temuan utama. Pertama, terkait dengan lanskap emosi dalam musik populer, dan kedua, mengenai efektivitas metodologi Natural Language Processing (NLP) dalam memetakannya. Secara tematik, analisis ini mengonfirmasi dominasi emosi positif dalam lirik lagu populer, dengan emosi 'romantis' (36,2%) dan 'bahagia' (26,2%) menjadi tema yang paling banyak dieksplorasi. Sementara itu, ekspresi 'kesedihan' (16,3%) juga muncul cukup signifikan, menunjukkan bahwa emosi negatif masih memiliki ruang dalam narasi musik populer. Emosi 'kemarahan' (4,4%), di sisi lain, tercatat sebagai yang paling jarang ditemukan, menciptakan ketidakseimbangan kelas yang signifikan. Secara metodologis, penggunaan pendekatan berbasis *TF-IDF* dan leksikon emosi kustom yang dikombinasikan dengan model *Random Forest* serta penyesuaian *class weighting* terbukti efektif, dengan capaian akurasi 83,03% dan *weighted average F1-score* sebesar 0,82. Hal ini menunjukkan bahwa pendekatan klasik dapat bekerja dengan baik dalam menangkap sinyal emosi dalam teks lirik dalam skala besar.

Namun, meskipun model ini menunjukkan kinerja yang solid, terdapat tantangan signifikan dalam mendeteksi emosi dengan frekuensi rendah, seperti 'kemarahan', yang hanya memiliki recall 42% dan *F1-score* 0,57. Kinerja yang kurang optimal pada emosi ini disebabkan oleh ketidakseimbangan distribusi data, keterbatasan leksikon dalam menangkap konteks emosional secara mendalam, serta tumpang tindih leksikal antar kategori emosi. Penelitian ini tidak hanya berhasil menyajikan pemetaan emosi musikal yang berbasis data secara empiris, tetapi juga membangun fondasi yang kokoh dan dapat direplikasi untuk penelitian selanjutnya. Sebagai langkah lanjut, disarankan untuk melakukan beberapa pengembangan, antara lain: (1) Mengimplementasikan teknik penanganan ketidakseimbangan data yang lebih agresif, seperti *SMOTE* atau *cost-sensitive learning* dengan *focal loss* jika menggunakan model *deep learning*; (2) Mengeksplorasi fitur yang lebih kaya konteks, seperti penggunaan *embedding* dari model berbasis *transformer* seperti *BERT* atau *IndoBERT*, yang mampu menangkap hubungan kontekstual antar kata; (3) Mengembangkan atau mengadaptasi leksikon emosi yang lebih komprehensif dan tervalidasi untuk domain lirik lagu, termasuk penanganan negasi dan slang; (4) Melakukan validasi manual pada subset data untuk mengukur akurasi pelabelan leksikon; (5) Memperluas cakupan bahasa ke bahasa Indonesia atau multibahasa untuk meningkatkan generalisasi temuan; dan (6) Mengintegrasikan analisis emosi lirik

dengan fitur audio untuk aplikasi sistem rekomendasi musik yang lebih holistik. Dengan rekomendasi ini, diharapkan analisis emosi dalam lirik lagu dapat menangkap spektrum perasaan manusia secara lebih merata, sensitif, dan akurat, membuka peluang aplikasi yang lebih luas dalam berbagai domain, termasuk sistem rekomendasi musik yang lebih adaptif dan personal.

5. Ucapan Terima Kasih

Artikel ini disusun sebagai tugas akhir dari penyelesaian Program Studi Strata I Teknik Informatika di Universitas Kristen Satya Wacana. Penulis ingin menyampaikan rasa terima kasih kepada semua pihak yang telah terlibat dalam penyelesaian penelitian ini. Kepada dosen pembimbing yang telah memberikan saran, bimbingan, dan motivasi berharga sejak tahap perencanaan hingga penyusunan artikel. Juga kepada tim reviewer yang telah memberikan masukan selama proses revisi. Terakhir, apresiasi disampaikan kepada keluarga dan sahabat yang senantiasa mendampingi dan memberikan segala bentuk dukungan moral, spiritual, serta keterlibatannya dalam diskusi konstruktif yang memperkaya kualitas penelitian ini. Semoga karya ini dapat bermanfaat bagi pengembangan ilmu pengetahuan dan masyarakat luas.

6. Daftar Pustaka

- Abdillah, J., Asror, I., & Wibowo, Y. F. A. (2020). Emotion classification of song lyrics using Bidirectional LSTM method. *Resti Journal*, 4(4), 723–729. <https://doi.org/10.29207/resti.v4i4.2156>.
- Akaishi, J., Sakata, M., Yoshinaga, J., Nakano, M., Koshi, K., & Kiyota, K. (2022). Estimating the emotional information in Japanese songs using search engines. *Sensors*, 22(5), 1–11. <https://doi.org/10.3390/s22051800>.
- Ara, A., & Rekha, V. (2024). Enhancing music emotion classification using multi-feature approach. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 15(9), 794–803. <https://doi.org/10.14569/IJACSA.2024.0150981>.
- Ardianti, I. A. P. G., Andriyani, A. A. A. D., & Swabawa, I. B. M. S. (2023). Discourse analysis on Jiwa yang Bersedih song lyrics. *Lire Journal (Journal of Linguistics and Literature)*, 7(2), 290–300. <https://doi.org/10.33019/lire.v7i2.227>.
- Bužić, D., & Dobša, J. (2025). Emotions in Eurovision Song Contest lyrics. *2025 MIPRO 48th ICT and Electronics Convention*, 137–142. <https://doi.org/10.1109/MIPRO65660.2025.11131930>.
- Chowdary, P., Singh, B., Agarwal, R., & Alluri, V. (2024). Lyrically speaking: Exploring the link between lyrical emotions, themes and depression risk. *25th International Society for Music Information Retrieval Conference*. <https://doi.org/10.48550/arXiv.2408.15575>.
- Czarnek, G., & Stillwell, D. (2022). Two is better than one: Using a single emotion lexicon can lead to unreliable conclusions. *PLOS One*, 17(10), 1–22. <https://doi.org/10.1371/journal.pone.0275910>.
- Gayatri, N. S. A., & Zulfiningrum, R. (2025). Kajian semiotika: Urgensi dukungan sosial dalam lirik lagu "Stay Alive" karya Bangtan Sonyeondan. *Jurnal Komunikasi Universitas Garut: Hasil Pemikiran dan Penelitian*, 11(1), 250–268. <https://doi.org/10.52434/jk.v11i1.41490>.

- Hutabarat, S. W., Helvira, J., Harahap, N. I., Manalu, N., Pasya, M. R., Hutaauruk, R. A. R., & Amalia, N. (2025). Afiksasi dalam lirik lagu: Studi kasus album *Short N' Sweet* karya Sabrina Carpenter. *Jurnal Sistem Informasi, Teknik Komputer dan Teknologi Pendidikan*, 5(1), 27–34. <https://doi.org/10.55338/justikpen.v5i1.154>.
- Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, 6, 100059. <https://doi.org/10.1016/j.nlp.2024.100059>.
- Jo, W., & Kim, M. J. (2022). Tracking emotions from song lyrics: Analyzing 30 years of K-Pop hits. *Emotion*, 23(6), 1658–1669. <https://doi.org/10.1037/emo0001185>.
- Mishra, N., & Chingre, S. S. (2024). Enhanced mood and theme recognition in music using lyrical sentiment analysis. *International Journal for Research*, 12(11), 2437–2442. <https://doi.org/10.22214/ijraset.2024.65624>.
- Mohammad, S. M. (2023). Best practices in the creation and use of emotion lexicons. *Findings of the Association for Computational Linguistics: EACL*, 1825–1836. <https://doi.org/10.18653/v1/2023.findings-eacl.136>.
- Ngo, G., Beard, R., & Chandra, R. (2022). Evolutionary bagging for ensemble learning. *Neurocomputing*, 510, 1–14. <https://doi.org/10.1016/j.neucom.2022.08.055>.
- Pratama, I. R., Cahyana, Y., & Wahiddin, D. (2025). Model machine learning untuk analisis sentimen masyarakat terhadap kenaikan PPN di media sosial X. *Bulletin of Computer Science Research*, 5(4), 277–286. <https://doi.org/10.47065/bulletincsr.v5i4.523>.
- Saptono, H., Achmar, Y. F., Hadi, H. S., Shiroth, S. F., Aria, L. R. P., & Alfarizqi, M. M. (2025). Implementasi deteksi intrusi aplikasi web berbasis supervised machine learning: Studi kasus LMS STT Terpadu Nurul Fikri. *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(3), 888–902. <https://doi.org/10.51454/decode.v5i3.1313>.
- Simbolon, L., Suhadi, J., & Pratiwy, D. (2025). Exploring the experiential meaning of song lyrics: A systematic literature review. *SIBATIK Journal: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan*, 4(11), 3661–3672. <https://doi.org/10.54443/sibatik.v4i11.3624>.
- Sutriawan, Rustad, S., Shidik, G. F., & Pujiono. (2025). Performance evaluation of text embedding models for ambiguity classification in Indonesian news corpus: A comparative study of TF-IDF, Word2Vec, FastText, BERT, and GPT. *International Information and Engineering Technology Association*, 30(6), 1469–1482. <https://doi.org/10.18280/isi.300606>.
- Yusupov, K., Muminov, I., Islam, M. R., & Sahlabadi, M. (2024). Comparative analysis of machine learning and deep learning models for email spam classification using TF-IDF and word embedding techniques. *International Conference on Broadband and Wireless Computing, Communication and Applications*. <https://doi.org/10.1007/978-3-031-76452-3>.