

IMPLEMENTATION OF TEXT MINING ON SONG LYRICS FOR SONG CLASSIFICATION BASED ON EMOTION USING WEBSITE-BASED LOGISTIC REGRESSION

Swahesti Puspita Rahayu^{*1}, Lasmedi Afuan², Galih Arditiya Yunindar³

^{1,2,3}Informatics, Engineering Faculty, Universitas Jenderal Soedirman, Indonesia

Email: ¹swahesti.rahayu@unsoed.ac.id, ²lasmedi.afuan@unsoed.ac.id, ³galih.yunindar@mhs.unsoed.ac

(Article received: February 12, 2024; Revision: February 13, 2024; published: February 20, 2025)

Abstract

Music has become an essential medium for expressing emotions and enriching human social experiences. However, the manual interpretation of emotions in song lyrics is often inaccurate and time-consuming, especially for complex or ambiguous lyrics. This creates a need for an automated system that can improve the accuracy and efficiency of emotion classification in song lyrics. Various algorithms, such as K-Nearest Neighbor (K-NN), Naive Bayes Classifier, and Support Vector Machine (SVM), have been applied for emotion classification in song lyrics. Previous research has shown that SVM combined with Particle Swarm Optimization (PSO) achieves an accuracy of up to 90%, while K-NN with feature selection produces the highest f-measure of 66.93%, and Naive Bayes achieves an accuracy of up to 45%. In this study, the Logistic Regression algorithm, supported by the Term Frequency-Inverse Document Frequency (TF-IDF) method, is applied to enhance the accuracy of emotion classification. Evaluation results indicate that the model with figurative language transformation achieves a higher accuracy (93.52%) compared to the model without figurative language transformation (92.31%), demonstrating that figurative language contributes to the richness of emotional expression recognized by the model. This model shows competitive results and can be compared to SVM using PSO while providing better performance than K-NN and Naive Bayes. The system implementation is web-based using the Streamlit framework, allowing users to input lyrics and obtain interactive emotion predictions. This research contributes to the analysis of music emotions and offers an efficient and more accessible alternative for emotion classification in song lyrics.

Keywords: Emotion, Logistic Regression, Text Mining, TF-IDF.

IMPLEMENTASI TEXT MINING PADA LIRIK LAGU UNTUK KLASIFIKASI LAGU BERDASARKAN EMOSI MENGGUNAKAN LOGISTIC REGRESSION BERBASIS WEBSITE

Abstrak

Musik telah menjadi media penting dalam mengekspresikan emosi dan memperkaya pengalaman sosial manusia. Namun, proses interpretasi emosi pada lirik lagu secara manual sering kali tidak akurat dan membutuhkan waktu, terutama pada lirik yang kompleks atau ambigu. Hal ini menimbulkan kebutuhan akan sistem otomatis yang dapat meningkatkan akurasi dan efisiensi dalam klasifikasi emosi lirik lagu. Berbagai algoritma, seperti K-Nearest Neighbor (K-NN), Naive Bayes Classifier, dan Support Vector Machine (SVM), telah diterapkan untuk klasifikasi emosi pada lirik lagu. Penelitian sebelumnya menunjukkan bahwa SVM yang dikombinasikan dengan Particle Swarm Optimization (PSO) mencapai akurasi hingga 90%, sementara K-NN dengan seleksi fitur menghasilkan f-measure tertinggi sebesar 66,93%, dan Naive Bayes mencapai akurasi hingga 45%. Dalam penelitian ini, algoritma Logistic Regression yang didukung oleh metode Term Frequency-Inverse Document Frequency (TF-IDF) Hasil evaluasi menunjukkan bahwa model dengan transformasi majas memiliki akurasi yang lebih tinggi (93.52%) dibandingkan dengan model tanpa transformasi majas (92.31%), menunjukkan bahwa majas memberikan kontribusi terhadap kekayaan ekspresi emosional yang dapat dikenali oleh model. Model ini menunjukkan hasil yang kompetitif dan dapat dibandingkan dengan SVM yang menggunakan PSO, sekaligus memberikan kinerja yang lebih baik dibandingkan K-NN dan Naive Bayes. Implementasi sistem ini berbasis website menggunakan framework Streamlit, memungkinkan pengguna memasukkan lirik dan mendapatkan prediksi emosi secara interaktif. Penelitian ini memberikan kontribusi dalam analisis emosi musik, serta menawarkan alternatif yang efisien dan aksesibilitas yang lebih luas untuk klasifikasi emosi dalam lirik lagu.

Kata kunci: Emosi, Logistic Regression, Text Mining, TF-IDF.

1. PENDAHULUAN

Hakikat musik pada dasarnya memang tidak bisa terpisahkan dari kehidupan manusia, musik sudah menjadi bagian dari peradaban manusia dari sejak lama hingga saat ini. Selain itu musik juga memainkan peran penting dalam mengekspresikan emosi dan meningkatkan keterampilan sosial [1], [2]. Sebagai media ekspresi, musik tidak hanya disampaikan melalui melodi dan ritme, tetapi juga melalui lirik yang sering kali menyampaikan makna mendalam dan mampu menyentuh perasaan pendengarnya. Hal ini membuat musik dan lirik menjadi objek kajian menarik dalam studi emosi dan perilaku manusia [3].

Lirik merupakan media ekspresi dalam sebuah musik, yang dimana dalam mengekspresikannya penyair menggunakan kekreativitasannya untuk memainkan kata, sehingga lirik lagu di dalamnya memiliki daya tarik tersendiri [4]. Namun, hal ini juga menciptakan tantangan karena lirik yang ambigu, metaforis, atau menggunakan bahasa kiasan sering kali sulit untuk diinterpretasi [5]. Selain itu, lirik lagu yang kompleks dapat mengandung berbagai emosi yang saling tumpang tindih, sehingga menambah tingkat kesulitan dalam proses klasifikasi emosi [6]. Sebagai contoh, lirik yang terlihat sederhana mungkin menyimpan makna yang mendalam dan kompleks yang hanya bisa dipahami melalui konteks dan analisis lebih lanjut. Ketidakpastian ini membuat banyak model klasifikasi emosi mengalami penurunan akurasi ketika dihadapkan pada *dataset* yang heterogen. Dengan meningkatnya jumlah lirik yang dipublikasikan setiap tahun, proses manual untuk memahami emosi menjadi tidak efisien dan membutuhkan solusi berbasis teknologi yang lebih efektif.

Beberapa penelitian telah dilakukan dalam klasifikasi emosi berbasis text mining dengan berbagai pendekatan algoritma, seperti *K-Nearest Neighbor* (K-NN), *Naïve Bayes Classifier*, dan *Support Vector Machine* (SVM) [7], [8], [9], [10], [11], [12]. Namun, meskipun metode-metode tersebut telah menunjukkan keberhasilan dalam klasifikasi emosi, masih terdapat tantangan yang belum sepenuhnya teratasi, terutama dalam menangani gaya bahasa figuratif dalam lirik lagu. Lirik lagu sering kali menggunakan ekspresi metaforis atau majas yang dapat memengaruhi interpretasi emosi jika tidak diproses dengan baik. Kekurangan ini membuka peluang untuk memperbaiki pendekatan preprocessing dalam text mining sehingga dapat mengakomodasi perubahan gaya bahasa dan meningkatkan akurasi model klasifikasi emosi.

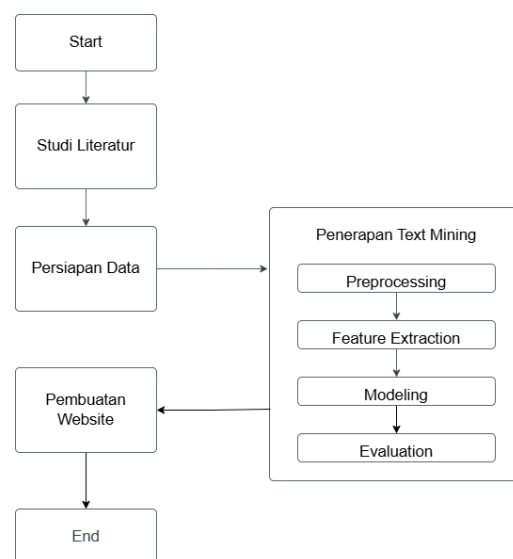
Penggunaan algoritma *Logistic Regression* dalam penelitian berbasis text mining sering kali dipilih karena keunggulannya dalam efisiensi dan interpretasi hasil. Seperti yang dijelaskan oleh [13], [14], [15], [16], [17], *logistic regression* memiliki

kemampuan untuk mengidentifikasi hubungan antara variabel independen dan hasil klasifikasi dengan cara yang transparan dan mudah dipahami. Keunggulan ini menjadi penting dalam penelitian berbasis lirik lagu, dimana pemahaman terhadap hubungan antara fitur (seperti kata-kata dalam lirik) dan label kelas (kategori emosi) dapat memberikan wawasan tambahan bagi pengguna. Selain itu, *logistic regression* lebih cepat dan membutuhkan sumber daya komputasi yang lebih sedikit dibandingkan dengan *Support Vector Machine* (SVM), yang sering memerlukan tuning parameter dan biaya komputasi yang lebih tinggi untuk menghasilkan hasil optimal. Dengan mempertimbangkan efisiensi dan kesesuaian dengan *dataset* berukuran kecil hingga menengah, algoritma *Logistic Regression* menjadi pilihan yang tepat untuk penelitian ini.

Berdasarkan latar belakang permasalahan tersebut, maka akan dilakukan sebuah penelitian dengan judul “Implementasi *Text Mining* Pada Lirik Lagu Untuk Klasifikasi Lagu Berdasarkan Emosi Menggunakan *Logistic Regression* Berbasis *Website*” Penelitian ini bertujuan untuk mengembangkan sistem berbasis website yang dapat mengklasifikasikan emosi dalam lirik lagu secara otomatis. Dengan menggunakan framework *Streamlit*, aplikasi ini memungkinkan pengguna untuk memasukkan lirik lagu dan mendapatkan prediksi emosi secara interaktif.

2. METODE PENELITIAN

Penelitian diawali dengan *Studi Literatur* untuk memahami konsep dan referensi penelitian terdahulu. Selanjutnya, dilakukan pengumpulan dan persiapan data secara online, diikuti oleh penerapan *Text Mining*. Tahap akhir adalah pengembangan *website* untuk implementasi model. Diagram alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahap Penelitian

2.1. Studi Literatur

Tahap ini mencakup perumusan teori dasar, referensi kerja, dan metode klasifikasi teks melalui tinjauan pustaka. Literatur mengenai teknik *text mining*, algoritma *Logistic Regression*, dan visualisasi hasil dengan *Streamlit* dikumpulkan dari buku, jurnal, serta artikel terkait klasifikasi emosi dalam lirik lagu..

2.2. Persiapan Data

Tahap ini dimulai dengan pengumpulan data berupa lirik lagu yang telah diberi label emosi, diperoleh dari *Kaggle* dan *Google* (2000-2023). Selanjutnya, data dipersiapkan dengan mengimpor *dataset*, memeriksa struktur dan tipe data, mengecek nilai *null*, serta melakukan analisis statistik deskriptif sebelum diproses lebih lanjut dengan *Text Mining*.

2.3. Text Mining

Text Mining adalah teknik untuk mengekstraksi informasi dari data teks menggunakan metode *machine learning*, *data mining*, dan *natural language processing*. [18]. Proses ini mencakup beberapa tahapan utama:

a. Text Preprocessing

Pada tahapan *preprocessing* merupakan kegiatan untuk Menyiapkan data mentah dengan menghapus elemen tidak terstruktur atau mengubahnya menjadi format yang lebih mudah diproses. [12].

b. Feature Selection

Setelah *preprocessing*, langkah berikutnya adalah Mengekstraksi fitur dari teks yang telah dibersihkan menggunakan teknik TF-IDF agar dapat diolah oleh model *machine learning*.

c. Modeling

Pada tahap ini, model klasifikasi dibangun menggunakan dengan algoritma *Logistic Regression Multinomial*, yang dipilih karena efisiensinya dalam menangani klasifikasi *multi-kelas*.

d. Evaluation

Setelah model dilatih, langkah selanjutnya adalah Mengukur kinerja model dengan menguji hasil klasifikasi menggunakan data uji untuk memastikan akurasi dan efektivitasnya.

2.4. TF-IDF

TF-IDF (*Term Frequency – Inverse Document Frequency*) adalah metode pembobotan kata dengan menentukan frekuensi kemunculan kata dalam sebuah dokumen. TF-IDF melakukan pembobotan kata (*term*) pada setiap dokumen dengan memperhatikan frekuensi atau seringnya kata tersebut muncul dalam sebuah dokumen. TF-IDF melakukan pembobotan kata (*term*) pada setiap dokumen dengan memperhatikan frekuensi atau seringnya kata tersebut muncul dalam sebuah

dokumen [16]. Perhitungan TF-IDF menggunakan persamaan (1)-(3).

$$tf_{t,d} = \frac{f_{t,d}}{n_d} \quad (1)$$

$$df_t = \ln \frac{N+1}{df_t+1} + 1 \quad (2)$$

$$W_{t,d} = tf_{t,d} \cdot idf_{t,d} \quad (3)$$

2.5. Logistic Regression

Logistic Regression adalah suatu metode yang dapat digunakan untuk mencari hubungan antara variabel respon yang bersifat *dichotomus* (skala nominal/ordinal dengan dua kategori) dengan satu atau lebih variabel prediktor berskala kategori atau kontinu. Model *Logistic Regression* terdiri dari *Logistic Regression* dengan *respon biner*, *ordinal*, dan *multinomial* [9]. Dalam penelitian ini akan menggunakan model *Logistic Regression Multinomial*.

Logistic Regression Multinomial merupakan *Logistic Regression* dengan variabel dependen (Y) mempunyai skala yang bersifat *polychotomus* atau *multinomial* yaitu skala dengan kategori lebih dari dua. Saat menghitung dengan metode ini, pertama kali yang dilakukan adalah menginisialisasi bobot dan bias, kemudian melakukan perhitungan probabilitas menggunakan *Softmax* dengan persamaan (4)-(6).

$$\pi_0(x) = \frac{1}{1+e^{g_1(x)}+e^{g_2(x)}} \quad (4)$$

$$\pi_1(x) = \frac{e^{g_1(x)}}{1+e^{g_1(x)}+e^{g_2(x)}} \quad (5)$$

$$\pi_2(x) = \frac{1e^{g_2(x)}}{1+e^{g_1(x)}+e^{g_2(x)}} \quad (6)$$

Setelah menghitung *softmax*, kemudian menghitung *Loss* dengan Cross-Entropy dengan persamaan (7).

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(\pi_j(x_i)) \quad (7)$$

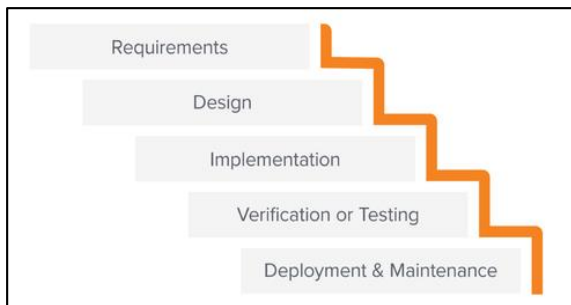
Selama proses pelatihan, model akan mengalami pembaruan parameter menggunakan *gradient descent* hingga memenuhi kriteria konvergensi, seperti nilai *loss* yang stabil atau tercapainya jumlah *epoch* yang telah ditentukan. Rumus *gradient descent* dapat dilihat pada persamaan (8).

$$\beta_j^{(t+1)} = \beta_j^{(t)} - \eta \cdot \frac{\partial Loss}{\partial \beta_j} \quad (8)$$

2.6. Pembuatan Website

Setelah model dikembangkan dan dievaluasi, langkah selanjutnya adalah membangun *website* agar pengguna dapat mengakses model dengan mudah. Proses ini diawali dengan pembuatan *use case diagram* dan perancangan antarmuka, yang kemudian diimplementasikan menggunakan *framework Streamlit*. Pengujian dilakukan dengan metode *Blackbox Testing*.

Metode pengembangan yang digunakan adalah Waterfall, yang terdiri dari lima tahap: requirements, design, implementation, testing, dan maintenance [19]. Metode ini dipilih karena sifatnya yang sistematis, terstruktur, dan mudah dipahami, sehingga cocok untuk pengembangan sistem. Gambaran dari metode ini dapat dilihat pada gambar 2 [20].



Gambar 2. Metode Waterfall

a. Requirements

Menentukan kebutuhan sistem sangat penting karena sistem informasi merupakan bagian dari sistem yang lebih besar. Pembuatan sistem diawali dengan mengidentifikasi kebutuhan yang akan diterapkan dalam sistem informasi. [21].

b. Design

Tahap desain dilakukan sebelum coding untuk memberikan gambaran lengkap mengenai sistem yang akan dibangun, termasuk tampilan, arsitektur, serta kebutuhan perangkat keras dan lunak. [22].

c. Implementation

Pada tahap ini, seluruh elemen sistem disatukan, termasuk persiapan fasilitas, pelatihan pengguna, dan simulasi sebelum sistem digunakan. [23].

d. Testing

Pengujian software dilakukan untuk memastikan sistem berfungsi sesuai kebutuhan. Pengujian ini menggunakan metode blackbox untuk verifikasi dan validasi program [24]. Pada bagian evaluasi atau pengujian, dilakukan dengan menggunakan uji *blackbox*.

e. Maintenance

Setelah sistem diluncurkan, dilakukan pemeliharaan untuk memastikan kinerjanya tetap optimal dan dapat digunakan dalam jangka panjang. [25].

3. HASIL DAN PEMBAHASAN

3.1. Persiapan Data

Data yang digunakan dalam penelitian ini adalah data sekunder tentang lirik lagu Indonesia yang di ambil secara online dalam bentuk file JSON. Data lirik lagu disajikan dalam tabel 1.

Tabel 1. Data Lirik Lagu

Artist	Lyric	Title	Emotion
Acha Septriasa	Disini kau dan Aku\n Terbiasa Bersama\n Menjalan...	My Heart	Ketertarikan
Acha Septriasa	Embun Di Pagi Buta\nMenebarkan Bau Asa\nDetik ..	Sampai Menutup Mata	Kesedihan
Ada Band	[Verse 1]\nBagaimana mestinya\nMembuat mu jatuh...	Haruskah Kumati	Kesedihan
...
Last Child	Caraku 'tuk menjaga hatiku\nDengan selalu memori...	Hancur Lebih Dulu	Kemarahan

3.2. Text Mining

3.2.1. Preprocessing

Tahap *preprocessing* dimulai dengan mengubah teks majas menjadi bentuk literal agar lebih mudah dipahami oleh model. Hasil transformasi ini dapat dilihat pada Tabel 2.

Tabel 2. Transformasi Teks

Artist	Lyric	Title	Emotion
Acha Septriasa	Disini kau dan Aku\n Terbiasa Bersama\n Menjalan...	My Heart	Ketertarikan
Acha Septriasa	Embun Di Pagi Buta\nMenebarkan Bau Asa\nDetik ..	Sampai Menutup Mata	Kesedihan
Ada Band	[Verse 1]\nBagaimana mestinya\nMembuatmu jatuh...	Haruskah Kumati	Kesedihan
...
Last Child	Caraku 'tuk menjaga hatiku\nDengan selalu memori...	Hancur Lebih Dulu	Kemarahan

Tahap berikutnya adalah augmentasi data meliputi penggantian sinonim dan *back translation* untuk menambah variasi teks. Kata-kata diganti dengan sinonim, lalu diterjemahkan ke bahasa Inggris dan kembali ke bahasa Indonesia. Hasil dari augmentasi yang dapat dilihat pada Tabel 3.

Tabel 3. Hasil Augmentasi Data

No	lyric	emotion
1	disini kau dan aku\nterbiasa bersama\nmenjalan...	Ketertarikan
2	Di sini Anda dan saya terbiasa dengan kasih sa...	Ketertarikan
3	Di sini Anda dan saya terbiasa dengan kasih sa...	Ketertarikan
...
1232	Betapa saya 'untuk menjaga hati saya dengan te...	Kemarahan

Selanjutnya, Teks yang telah ditransformasi kemudian diproses dengan mengubahnya menjadi huruf kecil, menghapus karakter tidak relevan, serta melakukan tokenisasi agar hanya kata-kata bermakna yang digunakan. Hasil *preprocessing* dapat dilihat pada tabel 4.

Tabel 4. Hasil Preprocessing

Lyric	Emotion
['kau', 'terbiasa', 'menjalani', 'kasih', 'sayang', 'bahagia',...]	Ketertarikan
['terbiasa', 'kasih', 'sayang', 'hidup', 'bahagia',...]	Ketertarikan
['terbiasa', 'kasih', 'sayang', 'hidup', 'bahagia',...]	Ketertarikan
...	...
['betapa', 'menjaga', 'hati', 'memikirkan', 'manusia',...]	Kemarahan

3.2.2. Feature Extraction

Pada tahap ini, Ekstraksi fitur dilakukan menggunakan TF-IDF untuk mengubah teks lirik lagu menjadi representasi numerik. Teknik ini mengukur kepentingan suatu kata berdasarkan frekuensi kemunculannya dalam dokumen dan kebalikannya dalam keseluruhan koleksi dokumen. Hasil perhitungan TF-IDF dapat dilihat pada tabel 5.

Tabel 5. Hasil TF-IDF

Term	TF				IDF	Wdt = tf.idf		
	L1	...	L1232	df		L1	...	L1232
Abadi	0	...	0	33	4,59	0	...	0
Abaikan	0	...	0	7	6,04	0	...	0
Abu	0	...	0	2	7,02	0	...	0
...
Zatku	0	...	0	1	7,42	0	...	0

3.2.3. Modeling

Pada tahap *modeling*, dilakukan proses pembelajaran mesin menggunakan *Logistic Regression Multinomial* untuk memprediksi emosi dari lirik lagu. Model dilatih dengan data yang telah diproses, di mana bobot (β) dan bias (b) diinisialisasi secara acak dan diperbarui selama pelatihan untuk meningkatkan akurasi klasifikasi. Hasil bobot dan bias setelah iterasi pembaruan menggunakan metode *Gradient Descent* dapat dilihat pada tabel 6 dan 7.

Tabel 6. Hasil Bobot

Bobot Kata	Kemarahan	Kesedihan	Kesenangan	Ketertarikan
Abadi	-0.0175	-0.0259	-0.0343	0.0777
Abaikan	0.0069	-0.0100	-0.0006	0.0036
an				
Abu	-0.0007	-0.0046	-0.0012	0.0066
...
Zatku	-0.0001	0.0083	-0.0001	-0.0080

Tabel 7. Hasil Bias

Kelas	Nilai Bias
Kemarahan	-0.06243
Kesedihan	0.04576
Kesenangan	-0.06373
Ketertarikan	0.08040

Tahap selanjutnya menerapkan Fungsi aktivasi *Softmax* untuk menghitung probabilitas setiap kelas, memastikan total probabilitas berjumlah satu sehingga setiap data dapat dipetakan ke kelas emosi dengan tepat. Hasil perhitungan *softmax* dapat dilihat pada tabel 8.

Tabel 8. Hasil Softmax

No	Kemarahan	Kesedihan	Kesenangan	Ketertarikan
1	0,00000044	0,00011037	0,00001024	0,99987893
2	0,00001056	0,00253979	0,00001492	0,99743471
3	0,00006679	0,00517756	0,00006761	0,99468802
...
1232	0,99823184	0,00158062	0,00015156	0,00003596

3.2.4. Evaluation

Evaluasi awal dilakukan untuk memastikan model *Logistic Regression* dapat memprediksi emosi dari lirik lagu dengan akurasi yang memadai. Pengujian dilakukan dengan memberikan input lirik yang telah melalui *preprocessing*, termasuk penghapusan karakter tidak relevan, transformasi majas, dan konversi ke vektor TF-IDF. Model kemudian memprediksi emosi, seperti "kemarahan," dengan waktu eksekusi 0.0087 detik, menunjukkan efisiensi proses dalam menganalisis emosi lirik secara otomatis.

3.3. Pembuatan Website

3.3.1. Tahap Requirement

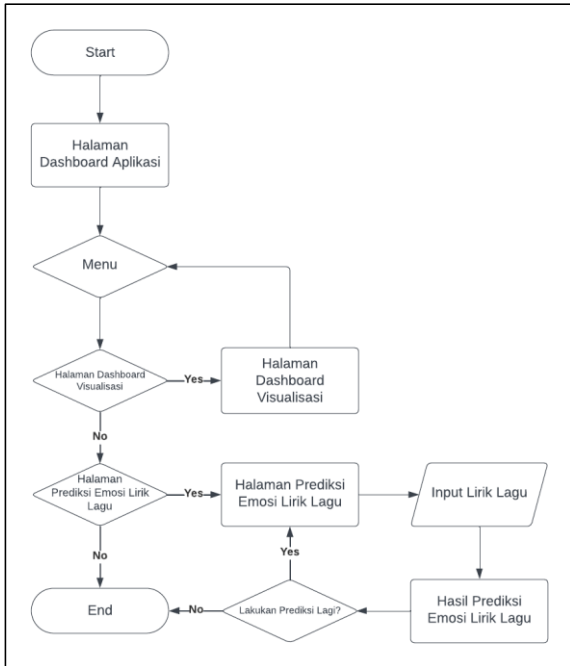
Tahap *requirements* merupakan langkah awal dalam perancangan sistem ini, dimana informasi mengenai permasalahan dan kebutuhan pengguna dikumpulkan untuk mendapatkan gambaran yang jelas mengenai sistem yang akan dibangun. Tujuannya adalah agar sistem dapat dirancang sesuai dengan kebutuhan yang telah diidentifikasi.

3.3.2. Tahap Design

Tahap desain merupakan langkah kedua setelah pengumpulan kebutuhan pengguna, bertujuan untuk merancang sistem agar pengembangannya lebih terarah. Pada tahap ini, dibuat rancangan sistem dalam bentuk *Flowchart*, *Use Case Diagram*, dan desain *mockup aplikasi* untuk memberikan gambaran jelas tentang sistem yang akan dibangun.

3.3.2.1. Flowchart

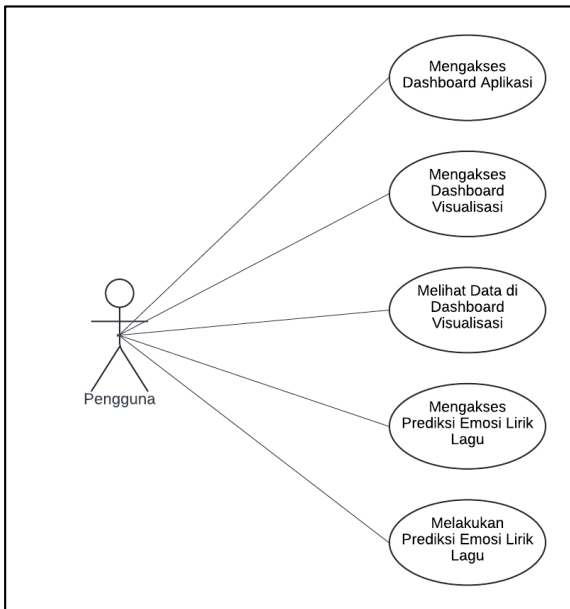
Flowchart menjelaskan bagaimana alur sistem ini berjalan baik dari proses awal sampai proses akhir. *Flowchart* dari aplikasi ini yang dapat dilihat pada Gambar 3.



Gambar 3. Flowchart

3.3.2.2. Use Case Diagram

Use Case Diagram merupakan gambaran arus informasi atau data yang diproses dari input menuju sebuah output tertentu. *Use Case Diagram* dari Aplikasi ini dapat dilihat pada Gambar 4.



Gambar 4. Use Case Diagram

3.3.2.3. Rancangan Antarmuka

Rancangan Antarmuka merupakan sebuah desain dasar bagaimana tampilan aplikasi yang akan dibuat.

3.3.3. Tahap Implementation

Aplikasi ini akan dibangun dengan menggunakan bahasa pemrograman *Python* dengan

framework Streamlit dan IDE *Visual Studio Code* yang terdiri dari tiga halaman utama.

1) Halaman *Dashboard Aplikasi*

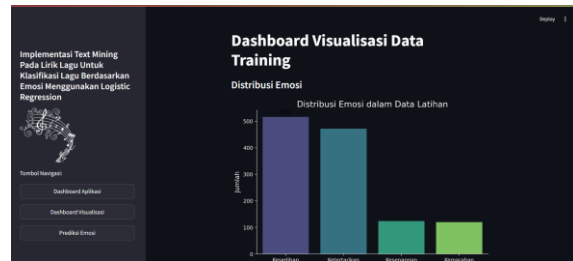
Halaman utama yang berisi informasi umum tentang aplikasi dan menu navigasi Tampilan dari halaman *Dashboard Aplikasi* dapat dilihat pada gambar 5.



Gambar 5. Halaman Dashboard Aplikasi

2) Halaman *Dashboard Visualisasi*

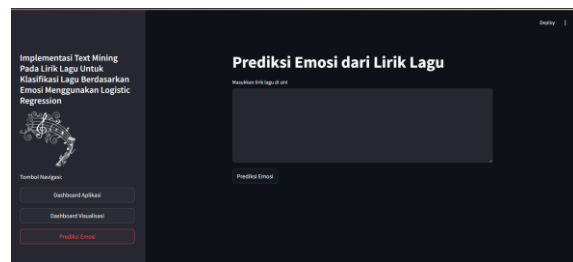
Menampilkan distribusi emosi dalam lirik lagu untuk memberikan wawasan kepada pengguna. Tampilan dari halaman *Dashboard Visualisasi* dapat dilihat pada gambar 6.



Gambar 6. Halaman Dashboard Visualisasi

3) Halaman *Prediksi Emosi Lirik Lagu*

Pengguna dapat memasukkan lirik lagu untuk dianalisis. Hasil prediksi emosi ditampilkan bersama dengan teks yang telah diproses. Tampilan dari halaman *Prediksi Emosi Lirik Lagu* dapat dilihat pada gambar 7.



Gambar 7. Halaman Prediksi Emosi Lirik Lagu.

3.3.4. Tahap Testing

Tahap *Testing* dilakukan menggunakan *Blackbox Testing* untuk menguji fungsionalitas sistem dan memastikan kesesuaiannya dengan rancangan. Pengujian ini bertujuan mengidentifikasi kekurangan agar dapat diperbaiki. Hasil pengujian disajikan pada Tabel 9.

Tabel 9. Blackbox Testing

No	Fitur	Test Case	Hasil Yang Diharapkan	Hasil
1	Dashboard Aplikasi	Mengklik “Dashboard Aplikasi” untuk mengarah ke halaman Dashboard Aplikasi. Pada saat membuka halaman Dashboard Aplikasi, menyajikan informasi mengenai website.	Website akan mengarahkan pengguna ke halaman Dashboard Aplikasi.	Valid
2	Dashboard Aplikasi	Mengklik “Dashboard Visualisasi” untuk mengarah ke halaman Dashboard Visualisasi. Pada halaman Dashboard Visualisasi, website akan menampilkan Visualisasi data, dapat melakukan pencarian data dan unduh sample data.	Website akan mengarahkan pengguna ke halaman Dashboard Visualisasi.	Valid
3	Dashboard Visualisasi	Mengklik “Prediksi Emosi Lirik Lagu” untuk mengarah ke halaman Prediksi Emosi Lirik Lagu. Pada halaman Prediksi Emosi Lirik Lagu, website akan menampilkan kolom input lirik lagu dan tombol prediksi emosi.	Website akan mengarahkan pengguna ke halaman Prediksi Emosi Lirik Lagu.	Valid
4	Dashboard Visualisasi	Melakukan input lirik lagu dan kemudian klik tombol Prediksi Emosi akan keluar output hasil prediksi emosi lirik lagu.	Website akan menampilkan data, melakukan pencarian data, dan unduh sample data.	Valid
5	Prediksi Emosi Lirik Lagu	Melakukan input lirik lagu dan kemudian klik tombol Prediksi Emosi akan keluar output hasil prediksi emosi lirik lagu.	Website akan menampilkan kolom input lirik lagu dan tombol prediksi emosi.	Valid
6	Prediksi Emosi Lirik Lagu	Melakukan input lirik lagu dan kemudian klik tombol Prediksi Emosi akan keluar output hasil prediksi emosi lirik lagu.	Website akan melakukan Prediksi Emosi Lirik Lagu dan mengeluarkan outputnya.	Valid
7	Prediksi Emosi Lirik Lagu	Melakukan input lirik lagu dan kemudian klik tombol Prediksi Emosi akan keluar output hasil prediksi emosi lirik lagu.	Website akan melakukan Prediksi Emosi Lirik Lagu dan mengeluarkan outputnya.	Valid

Selain melakukan pengujian dengan *blackbox*, Evaluasi model Logistic Regression dilakukan untuk memastikan keakuratan dan keandalan model dalam

mengklasifikasikan emosi pada lirik lagu. Perbandingan dilakukan antara dataset yang mengandung majas dan tanpa transformasi majas. Hasil evaluasi menunjukkan bahwa model dengan transformasi majas memiliki akurasi sebesar 93.52%, sementara model tanpa transformasi majas memiliki akurasi 92.31%. Dalam kategori kemarahan, precision pada kedua model mencapai 100%, tetapi recall lebih tinggi pada model tanpa transformasi majas. Untuk kategori kesedihan, hasilnya relatif konsisten dengan precision sekitar 90% dan recall 96% pada kedua model. Pada kategori kesenangan, precision tetap 100%, tetapi recall lebih tinggi pada model dengan transformasi majas. Sementara itu, kategori ketertarikan menunjukkan hasil yang hampir sama antara kedua model. Perbedaan ini menunjukkan bahwa keberadaan majas dalam lirik lagu dapat memengaruhi cara model menginterpretasikan emosi, dengan dampak yang berbeda pada precision dan recall di berbagai kategori. Selain itu, confusion matrix menunjukkan adanya perbedaan pola kesalahan klasifikasi antara kedua skenario, yang mengindikasikan bahwa majas berperan dalam kompleksitas interpretasi emosi oleh model.

3.3.5. Tahap Maintenance

Tahap *Maintenance* mencakup pengecekan rutin untuk memastikan aplikasi berfungsi dengan baik, pemantauan kinerja, perbaikan bug, serta pembaruan sistem guna meningkatkan fitur dan stabilitas aplikasi.

4. DISKUSI

Hasil penelitian ini menunjukkan bahwa penggunaan *Logistic Regression* dalam klasifikasi emosi lirik lagu memiliki peran penting dalam pengolahan data teks musik dan memberikan hasil yang kompetitif dibandingkan dengan metode sebelumnya, seperti *K-Nearest Neighbor* (K-NN), *Naïve Bayes Classifier*, dan *Support Vector Machine* (SVM). Model yang dikembangkan mampu mencapai akurasi 93.52% pada dataset dengan transformasi majas dan 92.31% pada dataset tanpa transformasi majas. Perbedaan akurasi ini menunjukkan bahwa majas memiliki pengaruh dalam klasifikasi emosi, di mana ekspresi figuratif dalam lirik dapat memperkuat atau melemahkan sinyal emosi yang dikenali oleh model.

Dibandingkan dengan penelitian sebelumnya, model ini menunjukkan peningkatan akurasi yang signifikan dengan pendekatan *preprocessing* yang lebih kompleks, termasuk transformasi majas dan seleksi fitur menggunakan TF-IDF. Studi ini juga menegaskan bahwa *Logistic Regression* lebih efisien dibandingkan dengan SVM dalam hal kebutuhan komputasi, namun tetap mampu memberikan hasil yang kompetitif.

Analisis confusion matrix menunjukkan bahwa model dengan transformasi majas memiliki keunggulan dalam mengklasifikasikan emosi tertentu, seperti kategori kesedihan dan ketertarikan, tetapi sedikit menurun pada kategori kemarahan dan kesenangan. Hal ini mengindikasikan bahwa model dapat lebih sensitif terhadap emosi yang memiliki ekspresi metaforis yang kuat. Selain itu, hasil evaluasi menunjukkan bahwa precision dan recall berbeda antar kategori, yang menunjukkan bahwa kompleksitas bahasa dalam lirik lagu memengaruhi kinerja model secara signifikan.

Dari segi implementasi, pengembangan sistem berbasis website dengan *framework Streamlit* memungkinkan pengguna untuk melakukan klasifikasi emosi lirik lagu secara interaktif dan efisien. Penggunaan *Blackbox Testing* dalam pengujian fungsionalitas sistem memastikan bahwa fitur utama aplikasi berjalan sesuai dengan spesifikasi yang dirancang.

Meskipun hasil penelitian ini cukup baik, terdapat beberapa keterbatasan yang perlu diperhatikan. Salah satunya adalah keterbatasan dataset yang digunakan, yang masih dapat diperluas untuk mencakup lebih banyak variasi genre dan bahasa. Selain itu, meskipun *preprocessing* majas telah diterapkan, tidak semua bentuk gaya bahasa dapat ditangani dengan sempurna. Oleh karena itu, penelitian lanjutan dapat mengembangkan metode lebih lanjut untuk menangani kompleksitas bahasa dalam lirik lagu, serta menguji pendekatan *deep learning* untuk meningkatkan akurasi model klasifikasi emosi.

5. KESIMPULAN

Keberadaan majas dalam lirik lagu terbukti memiliki dampak signifikan terhadap proses klasifikasi emosi. *Preprocessing* yang mencakup penghilangan atau pengelolaan majas merupakan langkah krusial untuk memastikan makna emosional tetap terjaga. Hasil evaluasi menunjukkan bahwa model dengan transformasi majas memiliki akurasi yang lebih tinggi (93.52%) dibandingkan dengan model tanpa transformasi majas (92.31%), yang menegaskan bahwa majas tidak hanya menambah kompleksitas pada teks, tetapi juga berkontribusi terhadap kekayaan ekspresi emosional yang dapat dikenali oleh model. Dengan semakin berkembangnya teknologi musik digital, diperlukan sistem otomatis yang mampu mengklasifikasikan emosi dalam lirik lagu secara akurat dan efisien, terutama dalam menangani kompleksitas bahasa figuratif. Analisis otomatis terhadap emosi dalam lirik lagu dapat dimanfaatkan untuk berbagai aplikasi, seperti sistem rekomendasi musik, analisis psikologi musik, hingga penelitian linguistik. Oleh karena itu, model klasifikasi dapat dikembangkan lebih lanjut dengan menerapkan metode *deep learning* seperti *Recurrent Neural Networks* (RNN) atau *Transformer* untuk meningkatkan pemahaman terhadap konteks

lirik yang lebih kompleks. Selain itu, pengujian dengan *dataset* yang lebih luas dan mencakup berbagai genre musik dapat memberikan wawasan lebih mendalam mengenai efektivitas model dalam berbagai kondisi linguistik.

DAFTAR PUSTAKA

- [1] F. Novriadi, F. Mayar, and Desyandri, "Memperkenalkan Drama Musikal Untuk Membangun Kreativitas," *Journal Of Social Science Research*, vol. 3, pp. 5757–5768, 2023, [Online]. Available: <https://j-innovative.org/index.php/Innovative>
- [2] D. Setiaji, "Naturalistic: Jurnal Kajian Penelitian dan Pendidikan dan Pembelajaran," *Naturalistic: Jurnal Kajian Penelitian dan Pendidikan dan Pembelajaran*, vol. 5, no. 1, pp. 692–697, 2020, [Online]. Available: <https://doi.org/10.35568/naturalistic.v5i1.909>
- [3] P. N. Juslin and J. Sloboda, *Handbook of music and emotion: Theory, research, applications*. Oxford University Press, 2011. [Online]. Available: <https://books.google.co.id/books?id=ZtwUDAAAQBAJ>
- [4] S. I. A. Cahya and G. G. Sukendro, "Musik sebagai Media Komunikasi Ekspresi Cinta (Analisis Semiotika Lirik Lagu 'Rumah ke Rumah' Karya Hindia)," *Koneksi*, vol. 6, no. 2, pp. 246–254, 2022, [Online]. Available: <https://doi.org/10.24912/kn.v6i2.15565>
- [5] B. Nushantari, "Metafora Kematian Pada Antologi Puisi Nur Ein Duft Und Windeswehen," *Journal Universitas Negeri Surabaya*, vol. 9, no. 2, 2020, [Online]. Available: <https://doi.org/10.26740/ide.v9n2.p%25p>
- [6] D. Edmonds and J. Sedoc, "Multi-Emotion Classification for Song Lyrics," *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 221–235, Apr. 2021, [Online]. Available: <https://aclanthology.org/2021.wassa-1.24>
- [7] L. Afuan, M. Khanza, and A. Z. Hasyati, "Enhancing Sentiment Analysis Of The 2024 Indonesian Presidential Inauguration On X Using Smote-Optimized Naive Bayes Classifier," *Jurnal Teknik Informatika (JUTIF)*, vol. 6, no. 1, pp. 325–333, 2025, doi: 10.52436/1.jutif.2025.6.1.4290.
- [8] L. Afuan and R. R. Isnanto, "Enhanced Fall Detection using Optimized Random Forest Classifier on Wearable Sensor Data," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 213–224, 2024, doi: 10.47738/jads.v6i1.498.

- [9] R. E. Fitri, "Analisis Regresi Logistik Biner Terhadap Data Indeks Kedalaman Kemiskinan Di Indonesia Tahun 2020," *Fakultas Matematika Dan Ilmu Pengetahuan Alam*, 2022, [Online]. Available: <http://digilib.unila.ac.id/63363/>
- [10] B. Ramadhani, R. R. Suryono, and K. Kunci, "JURNAL MEDIA INFORMATIKA BUDIDARMA Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse," 2024, doi: 10.30865/mib.v8i2.7458.
- [11] Y. T. Handika, S. Defit, and G. W. Nurcahyo, "Text Mining Dalam Membandingkan Metode Naïve Bayes Dengan C.45 Dalam Mengidentifikasi Berita Hoax Pada Media Sosial," *Rang Teknik Journal*, vol. 5, no. 1, pp. 116–123, Jan. 2022, doi: 10.31869/rtj.v5i1.2855.
- [12] A. Mustofa and R. Novita, "Klasifikasi Sentimen Masyarakat Terhadap Pemberlakuan Pembatasan Kegiatan Masyarakat Menggunakan Text Mining Pada Twitter," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, Jun. 2022, doi: 10.47065/bits.v4i1.1628.
- [13] N. C. Ramadani, "Analisis Sentimen Untuk Mengukur Ulasan Pengguna Aplikasi Mobile Legend Menggunakan Algoritma Naive Bayes, SVM, Random Fores, Decision Tree, dan Logistic Regression," *JSI : Jurnal Sistem Informasi (E-Journal)*, vol. 16, no. 1, pp. 123–138, 2024, [Online]. Available: <https://doi.org/10.18495/jsi.v16i1.152>
- [14] N. Silalahi and G. L. Ginting, "Rekomendasi Berita Berkaitan dengan Menerapkan Algoritma Text Mining dan TF-IDF," *Bulletin of Computer Science Research*, vol. 3, no. 4, pp. 276–282, Jun. 2023, doi: 10.47065/bulletincsr.v3i4.266.
- [15] W. Hermanto, B. Irawan, and C. Setianingsih, "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 Optimization," 2021. [Online]. Available: <https://lirik.kapanlagi.com/>
- [16] N. Hikmah, D. Ariyanti, and F. A. Pratama, "Implementasi Chatbot Sebagai Virtual Assistant di Universitas Panca Marga Probolinggo menggunakan Metode TF-IDF," *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 4, no. 2, pp. 133–148, Aug. 2022, doi: 10.35746/jtim.v4i2.225.
- [17] H. Dhery, A. Assyam, and F. N. Hasan, "Analisis Sentimen Twitter Terhadap Perpindahan Ibu Kota Negara Ke IKN Nusantara Menggunakan Orange Data Mining," *Media Online*, vol. 4, no. 1, pp. 341–349, 2023, doi: 10.30865/klik.v4i1.957.
- [18] T. Ridwansyah, "KLIK: Kajian Ilmiah Informatika dan Komputer Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *Media Online*, vol. 2, no. 5, pp. 178–185, 2022, [Online]. Available: <https://djournals.com/klik>
- [19] R. Rosdani and R. R. Waliansyah, "Sistem Informasi Pendataan Mahasiswa Magang Di Upt-Tik Universitas Pgrri Semarang," in *Proceeding Science and Engineering National Seminar*, 2020, pp. 66–71. [Online]. Available: <https://conference.upgris.ac.id/index.php/sens/article/view/1324>
- [20] A. Yusnita, E. Yulsilviana, and D. Sulissetiyo, "Sistem Informasi Pendataan Siswa Tidak Mampu Dan Putus Sekolah Pada Dinas Pendidikan Dan Kebudayaan Provinsi Kalimantan Timur Berbasis Web," *Jurnal Informatika Wicida*, vol. 8, no. 2, pp. 60–68, 2019, [Online]. Available: <https://jurnal.wicida.ac.id/index.php/informatika/article/view/1229>
- [21] T. Anjani, I. Mutia, and others, "Perancangan Aplikasi Peminjaman Buku Perpustakaan Pada SDN Mekarjaya 11 Depok," *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, vol. 5, no. 1, pp. 101–107, 2020, [Online]. Available: <https://www.academia.edu/download/68167767/3248.pdf>
- [22] J. Jamaludin and R. Romindo, "Sistem Informasi Akademik Berbasis Web Menggunakan Metode Waterfall Pada SMA Kemala Bhayangkari I Medan," in *Semantika (Seminar Nasional Teknik Informatika)*, 2019, pp. 17–27. [Online]. Available: <https://semantika.polgan.ac.id/index.php/Semantika/article/view/54>
- [23] I. M. S. Ardana, "Penguujian Software Menggunakan Metode Boundary Value Analysis dan Decision Table Testing," *Jurnal Teknologi Informasi ESIT*, vol. 14, no. 11, 2019, Accessed: Nov. 02, 2024. [Online]. Available: <https://www.researchgate.net/>
- [24] D. Rahmawati, A. S. Prabowo, and R. Purwanto, "Implementasi Model Waterfall pada Pengembangan Sistem Informasi Monitoring Prestasi Mahasiswa," *Journal of Innovation Information Technology and Application (JINITA)*, vol. 3, no. 1, pp. 82–93, Jun. 2021, doi: 10.35970/jinita.v3i1.678.

- [25] N. Hasanah and M. Nahrul Indriawan, "Rancangan Aplikasi Batam Travel Menggunakan Metode Software Development Life Cycle (SDLC)," *Conference On Management Business, Innovation, Education and Social Sciences*, vol. 1, no. 1, p. 925, 2021, [Online]. Available: <https://journal.uib.ac.id/index.php/combindes>.