



## **BERITA ACARA SIDANG PENDADARAN TUGAS AKHIR**

S/UBL/FTI/0203/VII/26

Pada hari ini, Senin 13 Juli 2026 telah dilaksanakan Ujian Sidang Pendadaran Tugas Akhir sebagai berikut:

Judul: ANALISIS SENTIMEN PADA LIRIK LAGU R&B DAN NEOSOUL MENGGUNAKAN  
TF-IDF DAN ALGORITMA NAIVE  
BAYES

Nama : Muhammad Andika Anggara Lukman Saudah  
NIM : 2211501248  
Dosen Pembimbing : Subandi, S.Kom., M.Kom.

Berdasarkan penilaian pada Presentasi + Demo, Penulisan, Penguasaan Materi, Penguasaan Program maka Mahasiswa tersebut di atas dinyatakan:

**LULUS**

dengan nilai angka : **80** huruf : **A-**

Mahasiswa tersebut di atas wajib menyerahkan hasil perbaikan tulisan Tugas Akhir dalam bentuk terjilid sesuai dengan Panduan Perbaikan Tugas Akhir, selambat-lambatnya Senin 27 Juli 2026.

### Panitia Penguji:

- |             |                                    |
|-------------|------------------------------------|
| 1 Ketua     | Dewi Kusumaningsih, S.Kom., M.Kom. |
| 2 Anggota   | Mufti, S.T., M.Kom.                |
| 3 Moderator | Subandi, S.Kom., M.Kom.            |

### Keterangan:

Nilai Huruf: A:85-100 A-:80-84,99 B+:75-79,99 B:70-74,99 B-:65-69,99 C:60-64,99 D:40-59,99  
E-:0-39,99



UNIVERSITAS BUDI LUHUR  
FAKULTAS TEKNOLOGI INFORMASI

Kartu Bimbingan Tugas Akhir

NIM: 2211501248

Nama: Muhammad Andika Anggara Lukman Saudah

Pembimbing: Subandi, S.Kom., M.Kom.

| No. | Tanggal    | Materi               |
|-----|------------|----------------------|
| 1   | 26-06-2026 | JUDUL                |
| 2   | 26-06-2026 | ABSTRAK              |
| 3   | 26-06-2026 | BAB I PENDAHULUAN    |
| 4   | 26-06-2026 | BAB 2 LANDASAN TEORI |
| 5   | 26-06-2026 | BAB 3                |
| 6   | 26-06-2026 | DEMO PROGRAM PART#1  |
| 7   | 26-06-2026 | BAB 4                |
| 8   | 26-06-2026 | DEMO PROGRAM PART#2  |
| 9   | 26-06-2026 | BAB 5                |



LEMBAR PENGESAHAN

|                       |                                                                                                       |
|-----------------------|-------------------------------------------------------------------------------------------------------|
| Nama                  | : Muhammad Andika Anggara Lukman Saudah                                                               |
| Nomor Induk Mahasiswa | : 2211501248                                                                                          |
| Program Studi         | : Teknik Informatika                                                                                  |
| Bidang Peminatan      | : Cyber Security                                                                                      |
| Jenjang Studi         | : Strata 1                                                                                            |
| Judul                 | : ANALISIS SENTIMEN PADA LIRIK LAGU R&B DAN NEOSOUL<br>MENGUNAKAN TF-IDF DAN ALGORITMA NAIVE<br>BAYES |



Laporan Tugas Akhir ini telah disetujui, disahkan dan direkam secara elektronik sehingga tidak memerlukan tanda tangan tim penguji.

Jakarta, Senin 13 Juli 2026

Tim Penguji:

|                     |                                      |
|---------------------|--------------------------------------|
| Ketua               | : Dewi Kusumaningsih, S.Kom., M.Kom. |
| Anggota             | : Mufti, S.T., M.Kom.                |
| Pembimbing          | : Subandi, S.Kom., M.Kom.            |
| Ketua Program Studi | : Dr. Indra, S.Kom., M.T.I.          |

**ANALISIS SENTIMEN PADA LIRIK LAGU *R&B* DAN *NEO-SOUL* MENGGUNAKAN TF-IDF DAN ALGORITMA NAIVE BAYES**

**TUGAS AKHIR**



**Oleh:**

**Muhammad Andika Anggara Lukman Saudah  
2211501248**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS BUDI LUHUR**

**JAKARTA  
2026**

# **ANALISIS SENTIMEN PADA LIRIK LAGU *R&B* DAN *NEO-SOUL* MENGGUNAKAN TF-IDF DAN ALGORITMA NAIVE BAYES**

**Diajukan Untuk Memenuhi Satu Persyaratan Memperoleh  
Gelar Sarjana Komputer (S.Kom)**



**Oleh:**

**Muhammad Andika Anggara Lukman Saudah  
2211501248**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS BUDI LUHUR**

**JAKARTA  
2026**

# LEMBAR PENGESAHAN



PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS BUDI LUHUR

## LEMBAR PENGESAHAN

Nama : Muhammad Andika Anggara Lukman Saudah  
Nomor Induk Mahasiswa : 2211501248  
Program Studi : Teknik Informatika  
Bidang Peminatan : Cyber Security  
Jenjang Studi : Strata 1  
Judul : ANALISIS SENTIMEN PADA LIRIK LAGU R&B DAN NEOSoul  
MENGUNAKAN TF-IDF DAN ALGORITMA NAIVE  
BAYES



Laporan Tugas Akhir ini telah disetujui, disahkan dan direkam secara elektronik sehingga tidak memerlukan tanda tangan tim penguji.

Jakarta, Senin 13 Juli 2026

Tim Penguji:

Ketua : Dewi Kusumaningsih, S.Kom., M.Kom.  
Anggota : Mufti, S.T., M.Kom.  
Pembimbing : Subandi, S.Kom., M.Kom.  
Ketua Program Studi : Dr. Indra, S.Kom., M.T.I.

## ABSTRAK

### ANALISIS SENTIMEN PADA LIRIK LAGU *R&B* DAN *NEO-SOUL* MENGUNAKAN TF-IDF DAN ALGORITMA NAIVE BAYES

Oleh: Muhammad Andika Anggara Lukman Saudah (2211501248)

Perkembangan industri musik digital menghasilkan lonjakan data teks tidak terstruktur berupa lirik lagu yang dikonsumsi secara masif oleh masyarakat. Lirik pada genre musik R&B dan Neo-Soul modern seringkali merekam dinamika emosi yang sangat kompleks dan menggunakan bahasa kiasan tingkat tinggi. Penelitian ini bertujuan untuk mengklasifikasikan polaritas sentimen pada lirik lagu genre tersebut ke dalam tiga kelas komputasional: Positif, Netral, dan Negatif, menggunakan pendekatan *Machine Learning*. Pemrosesan data dilakukan melalui serangkaian tahapan *Text Mining* yang meliputi *preprocessing* (*case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*), dilanjutkan dengan ekstraksi fitur menggunakan *Term Frequency - Inverse Document Frequency* (TF-IDF). Matriks bobot kata yang dihasilkan kemudian diklasifikasikan menggunakan algoritma *Multinomial Naive Bayes*. Evaluasi kinerja model dilakukan menggunakan metode *Confusion Matrix* dengan skenario pembagian data latih dan data uji sebesar 80:20 terhadap total 207 dokumen lirik lagu. Hasil pengujian menunjukkan bahwa model mampu mencapai nilai akurasi sebesar 57,14%. Analisis lebih lanjut mengungkap adanya ketimpangan prediksi (*skewness*) algoritma yang cenderung dominan memprediksi dokumen ke dalam kelas Positif. Hal ini disebabkan oleh tingginya frekuensi kemunculan *token* afeksi seperti kata "*love*" yang tetap dominan meskipun pada lirik bersentimen Negatif (patah hati/perpisahan). Keberadaan irisan leksikal ini bertindak sebagai *noise* yang mengelabui probabilitas mesin, sehingga penelitian ini memberikan wawasan baru mengenai tantangan klasifikasi pemrosesan bahasa alami (NLP) pada karya sastra musik berseni tinggi.

**Kata Kunci:** Analisis Sentimen, Lirik Lagu, R&B, Neo-Soul, TF-IDF, *Multinomial Naive Bayes*, *Confusion Matrix*.

## DAFTAR ISI

|                                                                    |      |
|--------------------------------------------------------------------|------|
| LEMBAR PENGESAHAN.....                                             | iii  |
| ABSTRAK .....                                                      | iv   |
| SURAT TIDAK PLAGIAT .....                                          | v    |
| SURAT AI GENERATIF .....                                           | vi   |
| KATA PENGANTAR.....                                                | vii  |
| DAFTAR SIMBOL.....                                                 | viii |
| DAFTAR ISI .....                                                   | ix   |
| DAFTAR TABEL .....                                                 | xii  |
| DAFTAR GAMBAR .....                                                | xiii |
| DAFTAR ALGORITMA .....                                             | xiv  |
| BAB I .....                                                        | 1    |
| 1.1 Latar Belakang.....                                            | 1    |
| 1.2 Perumusan Masalah .....                                        | 2    |
| 1.3 Batasan Masalah .....                                          | 2    |
| 1.4 Tujuan .....                                                   | 3    |
| 1.5 Manfaat .....                                                  | 3    |
| 1.6 Sistematika Penulisan .....                                    | 3    |
| BAB II.....                                                        | 5    |
| 2.1 <i>Natural Language Processing</i> (NLP).....                  | 5    |
| 2.2 <i>Text Mining</i> .....                                       | 5    |
| 2.3 <i>Knowledge Discovery in Databases</i> (KDD) .....            | 6    |
| 2.4 Analisis Sentimen.....                                         | 6    |
| 2.5 Ekstraksi Kata Kunci.....                                      | 7    |
| 2.6 Lirik Lagu Sebagai Objek Penelitian .....                      | 7    |
| 2.6.1 Pengertian Lirik Lagu.....                                   | 8    |
| 2.6.2 Genre <i>R&amp;B</i> .....                                   | 8    |
| 2.6.3 Genre <i>Neo-Soul</i> .....                                  | 9    |
| 2.6.4 Karakteristik Lirik <i>R&amp;B</i> dan <i>Neo-Soul</i> ..... | 9    |
| 2.7 <i>Pre-processing</i> .....                                    | 9    |
| 2.7.1 <i>Case Folding</i> .....                                    | 10   |
| 2.7.2 <i>Cleaning</i> .....                                        | 10   |
| 2.7.3 <i>Tokenizing</i> .....                                      | 10   |
| 2.7.4 <i>Stopword Removal</i> .....                                | 10   |
| 2.7.5 <i>Stemming</i> .....                                        | 11   |
| 2.8 <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF)..... | 11   |
| 2.9 Algoritma <i>Naïve Bayes</i> .....                             | 12   |
| 2.10 Evaluasi Model.....                                           | 13   |
| 2.10.1 <i>Confusion Matrix</i> .....                               | 13   |
| 2.10.2 <i>Accuracy</i> .....                                       | 13   |
| 2.10.3 <i>Precision</i> .....                                      | 14   |
| 2.10.4 <i>Recall</i> .....                                         | 14   |
| 2.10.5 <i>F1-Score</i> .....                                       | 14   |
| 2.11 Studi Literatur.....                                          | 15   |

|                                                                           |    |
|---------------------------------------------------------------------------|----|
| BAB III.....                                                              | 18 |
| 3.1 Metode Penelitian .....                                               | 18 |
| 3.1.1 Kerangka Penelitian .....                                           | 18 |
| 3.1.2 Arsitektur Sistem .....                                             | 19 |
| 3.2 Dataset dan Sumber Data Lirik Lagu .....                              | 19 |
| 3.3 Seleksi Data ( <i>Data Selection</i> ) .....                          | 20 |
| 3.4 Pra-Pemrosesan ( <i>Pre-Processing</i> ) .....                        | 21 |
| 3.4.1 <i>Case Folding</i> .....                                           | 21 |
| 3.4.2 Pembersihan ( <i>Cleaning</i> ) .....                               | 21 |
| 3.4.3 Tokenizing .....                                                    | 22 |
| 3.4.4 Penghapusan Kata Umum ( <i>Stopword Removal</i> ) .....             | 22 |
| 3.4.5 Stemming / Lemmatization .....                                      | 23 |
| 3.5 Ekstraksi Fitur Menggunakan TF-IDF .....                              | 23 |
| 3.6 Klasifikasi Sentimen Menggunakan Multinomial <i>Naive Bayes</i> ..... | 24 |
| 3.7 Rancangan Basis Data .....                                            | 24 |
| 3.7.1 <i>Logical Record Structure</i> (LRS) .....                         | 24 |
| 3.7.2 <i>Class Diagram</i> .....                                          | 25 |
| 3.7.3 Spesifikasi Basis Data .....                                        | 26 |
| 3.8 Rancangan Pengujian .....                                             | 28 |
| 3.9 Rancangan Menu .....                                                  | 29 |
| 3.10 Rancangan Layar .....                                                | 29 |
| 3.10.1 Rancangan Layar Login .....                                        | 30 |
| 3.10.2 Rancangan Layar Dashboard .....                                    | 30 |
| 3.10.3 Rancangan Layar Upload CSV .....                                   | 31 |
| 3.10.4 Rancangan Layar Data Lirik Asli ( <i>Raw Dataset</i> ) .....       | 31 |
| 3.10.5 Rancangan Layar <i>Pre-Processing</i> .....                        | 32 |
| 3.10.6 Rancangan Layar Ekstraksi Fitur (TF-IDF) .....                     | 32 |
| 3.10.7 Rancangan Layar Tabel Hasil Klasifikasi .....                      | 33 |
| 3.10.8 Rancangan Layar Visualisasi Distribusi Sentimen .....              | 33 |
| 3.10.9 Rancangan Layar Evaluasi Model .....                               | 34 |
| BAB IV HASIL DAN PEMBAHASAN .....                                         | 35 |
| 4.1 Lingkungan Percobaan .....                                            | 35 |
| 4.1.1 Spesifikasi Hardware .....                                          | 35 |
| 4.1.2 Spesifikasi Software .....                                          | 35 |
| 4.1.3 Deployment Diagram .....                                            | 35 |
| 4.2 Implementasi Sistem .....                                             | 36 |
| 4.2.1 Implementasi Dataset .....                                          | 36 |
| 4.2.2 Implementasi Preprocessing .....                                    | 37 |
| 4.2.3 Implementasi TF-IDF .....                                           | 40 |
| 4.2.4 Implementasi Multinomial Naive Bayes .....                          | 41 |
| 4.2.5 Implementasi Basis Data .....                                       | 44 |
| 4.3 Flowchart Sistem .....                                                | 45 |
| 4.3.1 Flowchart Keseluruhan Sistem .....                                  | 45 |
| 4.3.2 Flowchart Preprocessing .....                                       | 46 |
| 4.3.3 Flowchart Klasifikasi Sentimen .....                                | 47 |
| 4.4 Algoritma Sistem .....                                                | 48 |

|                                            |    |
|--------------------------------------------|----|
| 4.4.1 Algoritma Keseluruhan Sistem .....   | 48 |
| 4.4.2 Algoritma Preprocessing .....        | 49 |
| 4.4.3 Algoritma Klasifikasi Sentimen ..... | 49 |
| 4.5 Pengujian dan Evaluasi .....           | 50 |
| 4.5.1 Skenario Pengujian .....             | 50 |
| 4.5.2 Evaluasi Model Klasifikasi .....     | 50 |
| 4.5.3 Analisis Hasil Sentimen .....        | 52 |
| 4.5.4 Analisis Kata Kunci Dominan .....    | 52 |
| 4.6 Tampilan Layar Aplikasi .....          | 53 |
| 4.6.1 Login .....                          | 53 |
| 4.6.2 Dashboard .....                      | 53 |
| 4.6.3 Upload Dataset .....                 | 54 |
| 4.6.4 Dataset .....                        | 54 |
| 4.6.5 Preprocessing .....                  | 55 |
| 4.6.6 TF-IDF .....                         | 55 |
| 4.6.7 Hasil Klasifikasi .....              | 56 |
| 4.6.8 Distribusi Sentimen .....            | 56 |
| 4.6.9 Evaluasi Model .....                 | 57 |
| BAB V PENUTUP .....                        | 58 |
| 5.1 Kesimpulan .....                       | 58 |
| 5.2 Saran .....                            | 59 |
| DAFTAR PUSTAKA .....                       | 60 |
| LAMPIRAN .....                             | 62 |

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan pesat industri digital menghasilkan lonjakan *unstructured data* (data tidak terstruktur) berupa teks di berbagai platform media, salah satunya adalah lirik lagu. Dalam ruang lingkup *Natural Language Processing* (NLP) dan *text mining*, lirik lagu tidak sekadar susunan kata artistik, melainkan sebuah korpus data yang merekam dinamika psikologis, pergeseran nilai budaya, dan realita sosial masyarakat pendengarnya. Pada genre musik kontemporer seperti R&B dan *neo-soul*, struktur penulisan lirik memiliki karakteristik yang sangat unik dan sarat metafora. Narasi yang diangkat kini semakin kompleks dan personal; mulai dari keresahan menghadapi fase pendewasaan (*hustle culture*), pencarian ketenangan batin, hingga bentuk apresiasi dan kasih sayang yang mendalam terhadap sosok figur penenang, seperti keluarga atau ibu.

Meskipun lirik lagu genre ini dikonsumsi secara masif sebagai bentuk *coping mechanism* oleh masyarakat urban, pemahaman mengenai kecenderungan sentimen dan kata kunci yang paling mendominasi masih bersifat sangat subjektif. Di sisi lain, para produser dan musisi independen yang bergerak di kancah ini sering kali mengalami hambatan dalam memetakan *emotional positioning* dari sebuah karya. Penentuan arah narasi lagu saat ini mayoritas hanya mengandalkan intuisi artistik tanpa adanya dukungan data faktual mengenai kata atau narasi apa yang sebenarnya beresonansi kuat dengan pendengar modern.

Beberapa penelitian sebelumnya telah berupaya mengimplementasikan teknik *text mining* pada data lirik lagu. Penelitian yang dilakukan oleh (Sabda Husada & Yulianto, 2026) berfokus pada analisis lanskap emosi dalam lirik lagu menggunakan pendekatan NLP. Penelitian lain oleh (Kholida Zia Abidin et al., 2023) juga telah membuktikan efektivitas klasifikasi teks berskala besar pada lirik lagu daerah di Indonesia. Pendekatan serupa turut digunakan dalam perancangan sistem berbasis sentimen, di mana algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) terbukti akurat dalam mengekstraksi fitur teks lirik untuk memetakan emosi dan merekomendasikan penyesuaian *mood* (Syafaah & Lestandy, 2022). Selain itu, algoritma *Naive Bayes* juga telah tervalidasi sangat tangguh dalam membedah lirik lagu bertema khusus secara efisien (Ode et al., 2022).

Meskipun berbagai literatur tersebut telah membuktikan kehandalan komputasi teks pada lirik lagu, masih terdapat kesenjangan penelitian (*research gap*) yang dapat dikembangkan. Penelitian terdahulu rata-rata menggunakan dataset lirik yang bersifat umum atau *mainstream*, dibatasi pada klasifikasi regional geografis, atau masih memisahkan antara ekstraksi kata kunci dan analisis sentimennya. Belum ada penelitian komputasional yang memfokuskan penggalian kata kunci secara bersamaan dengan analisis sentimen yang secara spesifik menargetkan ekosistem lirik genre R&B dan neo-soul.

Untuk menjembatani kesenjangan tersebut, penelitian ini mengusulkan "Analisis Sentimen Pada Lirik Lagu R&B dan Neo-Soul Menggunakan TF-IDF

dan Algoritma Naive Bayes". Algoritma TF-IDF dipilih karena kehandalannya dalam memberikan bobot pada kata-kata spesifik yang jarang muncul secara global namun menjadi ciri khas inti dari sebuah teks, sehingga sangat cocok untuk mengekstraksi kata kunci dari gaya bahasa neo-soul. Sementara itu, Naive Bayes diimplementasikan sebagai model klasifikasi berbasis probabilitas yang efisien dalam menangani dataset teks berskala besar untuk memprediksi polaritas sentimen (positif, negatif, atau netral). Melalui pemodelan komputasi ini, diharapkan sentimen dan kata dominan lirik yang sebelumnya abstrak dapat divisualisasikan secara faktual, memberikan wawasan berbasis data (*data-driven*) yang objektif bagi perkembangan ekosistem musik independen.

## 1.2 Perumusan Masalah

Berdasarkan pemaparan latar belakang di atas, terdapat beberapa identifikasi masalah yang menjadi fokus utama dalam penyusunan Tugas Akhir ini. Adapun rumusan masalah tersebut adalah sebagai berikut:

1. Bagaimana mengimplementasikan algoritma TF-IDF untuk mengekstraksi fitur teks dan memetakan kata kunci utama yang merepresentasikan karakteristik narasi pada lirik lagu genre R&B dan Neo-Soul?
2. Bagaimana membangun model klasifikasi menggunakan algoritma *Naive Bayes* untuk mengukur polaritas sentimen lirik (positif, negatif, dan netral) guna melihat kecenderungan emosi pendengar modern secara objektif?
3. Bagaimana mengukur performa dan tingkat akurasi algoritma dalam memproses data teks sastra modern yang memiliki tingkat kompleksitas metafora tinggi?

## 1.3 Batasan Masalah

Mengingat luasnya cakupan permasalahan yang ada pada ekosistem musik dan pemrosesan bahasa alami, maka ruang lingkup penelitian ini perlu dibatasi agar pembahasan lebih terfokus, terarah, dan tidak menyimpang dari sasaran utama. Adapun batasan masalah yang ditetapkan dalam penelitian ini adalah sebagai berikut:

1. Dataset lirik lagu yang digunakan dibatasi pada genre R&B dan *Neo-Soul* dengan jumlah minimal 200 baris data teks.
2. Sistem menerapkan tahapan KDD yang mencakup tahapan pra-pemrosesan (*pre-processing*) seperti *cleaning*, *stemming*, dan *stopword removal*, serta transformasi data.
3. Algoritma yang digunakan berfokus pada pembobotan kata menggunakan TF-IDF dan algoritma klasifikasi *Naive Bayes*.
4. Program yang dihasilkan memiliki tampilan *Graphical User Interface* (GUI) berbasis *web* dan tidak disajikan dalam bentuk terminal atau *console*.
5. Program yang dirancang memuat fitur *import file dataset* menggunakan mode *browse* yang minimal mampu membaca satu jenis format data seperti *file .xls*, *.xlsx*, atau *.csv*.

## 1.4 Tujuan

Sejalan dengan rumusan masalah yang telah ditetapkan sebelumnya, penelitian ini dilaksanakan dengan beberapa tujuan utama. Adapun tujuan dari penelitian Tugas Akhir ini adalah:

1. Menerapkan metode *Text Mining* dengan pembobotan TF-IDF untuk melakukan ekstraksi dan identifikasi kata kunci dominan pada dataset lirik lagu.
2. Membangun sebuah sistem komputasi berbasis algoritma *Naive Bayes* yang mampu melakukan klasifikasi sentimen otomatis pada data tidak terstruktur (*unstructured data*).
3. Melakukan evaluasi terhadap kehandalan algoritma dalam memahami struktur bahasa urban dan metafora pada genre R&B melalui pengujian akurasi sistem.

## 1.5 Manfaat

Hasil dari penelitian yang dilakukan diharapkan tidak hanya memberikan kontribusi pada ranah akademis, tetapi juga memberikan dampak positif dan nilai guna bagi pihak-pihak terkait. Adapun manfaat yang diharapkan dari penelitian ini terbagi menjadi beberapa aspek, yaitu:

1. Memberikan perspektif baru dan validasi berbasis data mengenai kondisi psikologis serta pergeseran budaya yang tercermin dalam konsumsi musik sehari-hari, sehingga masyarakat dapat memahami tren emosi kolektif mereka secara lebih nyata.
2. Menjadi instrumen pendukung keputusan bagi musisi independen dan produser dalam merumuskan konsep karya yang lebih relevan dengan keresahan sosial pendengar, sehingga meningkatkan resonansi karya di pasar musik.
3. Memberikan kontribusi pada pengembangan studi *Natural Language Processing* (NLP), khususnya dalam pengolahan teks yang kaya akan bahasa kiasan dan idiom, guna memperkaya literatur teknis terkait tantangan pemrosesan bahasa alami di Indonesia.

## 1.6 Sistematika Penulisan

Penyusunan tugas akhir ini, disusun secara sistematis sehingga kaitan antara bab yang satu dengan bab yang lain dapat terlihat jelas. Oleh karena itu dalam 4 penulisan terbagi menjadi 5 (lima) bab, dimana setiap bab terdiri dari beberapa sub bab. Berikut ini adalah uraian singkat dari masing masing bab tersebut:

### BAB I

### PENDAHULUAN

Bab ini membahas tentang pendahuluan dari penelitian. Pendahuluan meliputi latar belakang pemilihan topik tugas akhir, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan tugas akhir ini.

### BAB II

### LANDASAN TEORI

Bab ini berisi tentang penjelasan serta pengertian-pengertian teori terkait, dan metode algoritma yang dipakai

dalam penelitian, serta literatur, referensi ilmiah yang digunakan sebagai penunjang dalam pembuatan aplikasi.

### **BAB III**

#### **METODOLOGI PENELITIAN**

Bab ini menjelaskan tentang metodologi yang diterapkan terkait permasalahan dan solusi dari aplikasi yang dikembangkan secara bertahap mulai dari data penelitian, penerapan metode yang digunakan, rancangan pengujian aplikasi yang diusulkan beserta cara kerjanya, rancangan layar, flowchart dan algoritma.

### **BAB IV**

#### **HASIL DAN PEMBAHASAN**

Bab ini menjelaskan tentang perangkat keras, perangkat lunak, sistem operasi yang digunakan, implementasi, mekanisme kerja dan aplikasi yang telah dibuat.

### **BAB V**

#### **PENUTUP**

Bab ini berisikan kesimpulan yang menjelaskan secara singkat mengenai hal yang dicapai dari program yang dikembangkan, serta saran-saran agar program ini dapat bekerja dengan baik.

## BAB II LANDASAN TEORI

### 2.1 *Natural Language Processing (NLP)*

*Natural Language Processing* (NLP) atau pemrosesan bahasa alami merupakan pendekatan komputasi yang digunakan untuk menganalisis dan memahami bahasa manusia, termasuk dalam pengenalan emosi pada sebuah teks komputasional. Pemanfaatan teknik NLP sangat relevan dalam pemrosesan data tekstual yang bersifat tidak terstruktur (*unstructured data*), seperti halnya dalam menganalisis sentimen pada lirik lagu (Manueke et al., 2023). Melalui disiplin ilmu ini, sistem komputasi dibekali kemampuan untuk menguraikan struktur linguistik dan mengekstraksi makna dari susunan kata, sehingga mesin atau komputer dapat mengenali ungkapan emosi yang terkandung di dalam teks lirik lagu (Sabda Husada & Yulianto, 2026).

Dalam konteks penelitian teks, karya musik sering kali memuat diksi kiasan, ambiguitas, serta ungkapan perasaan yang kompleks. Pengimplementasian teknik pemrosesan bahasa memfasilitasi transformasi teks lirik yang awalnya memiliki tingkat ambiguitas tinggi menjadi representasi data yang terukur dan objektif (Manueke et al., 2023). Dengan demikian, teks bahasa alami manusia tersebut siap untuk dieksekusi lebih lanjut ke dalam tahapan penambangan teks (*text mining*) dan klasifikasi algoritma (Sabda Husada & Yulianto, 2026).

### 2.2 *Text Mining*

Penambangan teks (*text mining*) merupakan suatu proses komputasi yang bertujuan untuk mengekstraksi informasi, pola tersembunyi, dan pengetahuan baru dari sekumpulan data teks yang bersifat tidak terstruktur (*unstructured data*). Berbeda dengan penambangan data (*data mining*) konvensional yang umumnya memproses angka pada basis data terstruktur, *text mining* secara spesifik difungsikan untuk menganalisis dokumen berbasis teks. Dalam berbagai penelitian terapan, implementasi *text mining* sangat krusial untuk melakukan analisis sentimen, di mana sistem komputasi ditugaskan untuk mengenali serta mengklasifikasikan kecenderungan polaritas teks secara otomatis (Tri Widodo & Setyawan, 2025).

Untuk dapat memproses data linguistik yang kompleks dan ambigu, penambangan teks terintegrasi secara erat dengan teknik *Natural Language Processing* (NLP) (Manueke et al., 2023). Integrasi ini memungkinkan sistem untuk melakukan pembersihan dan standarisasi teks mentah sebelum ditransformasikan menjadi representasi numerik. Pada konteks karya musik atau lirik lagu, penerapan *text mining* memfasilitasi algoritma komputer untuk memetakan emosi dominan di dalam lirik tersebut secara terukur (Sabda Husada & Yulianto, 2026). Lebih lanjut, selain berfungsi dalam pengklasifikasian sentimen, penambangan teks juga digunakan sebagai metode utama dalam proses ekstraksi kata kunci guna mengidentifikasi *term* dominan yang merepresentasikan gagasan utama di dalam sekumpulan dokumen (Odhianto et al., 2025).

### 2.3 Knowledge Discovery in Databases (KDD)

*Knowledge Discovery in Databases* (KDD) merupakan suatu kerangka kerja (*framework*) terstruktur yang digunakan untuk mengekstraksi informasi dan pengetahuan (*knowledge*) berharga dari sekumpulan data komputasional. Dalam konteks analisis data teks, KDD bertindak sebagai panduan metodologi sistematis yang memastikan bahwa proses penemuan pengetahuan berjalan secara runut, terukur, dan objektif. Penerapan kerangka kerja ini sangat krusial dalam penelitian penambangan teks (*text mining*), karena keseluruhan prosesnya merangkum siklus komputasi dari tahap inisiasi data mentah hingga tahap pengujian model klasifikasi sentimen secara utuh (Tri Widodo & Setyawan, 2025).

Secara arsitektur, tahapan KDD saling berkesinambungan dan mengalir secara sekuensial. Proses ini diawali dengan seleksi data (*data selection*) untuk mengumpulkan dan menentukan korpus teks yang valid dan relevan dengan parameter penelitian. Selanjutnya, data teks mentah tersebut dibersihkan dari berbagai ketidakteraturan leksikal melalui serangkaian teknik pra-pemrosesan (*pre-processing*) (Manueke et al., 2023). Setelah data teks menjadi bersih dan seragam, proses dilanjutkan dengan tahapan transformasi (*transformation*) untuk mengonversi fitur kata menjadi matriks vektor numerik.

Tahap inti komputasi kemudian dieksekusi pada fase penambangan data (*data mining*) dengan mengimplementasikan algoritma pembelajaran mesin (*machine learning*) untuk menemukan pola, klasifikasi polaritas emosi, atau ekstraksi topik dominan pada korpus lirik. Pada akhirnya, seluruh hasil klasifikasi tersebut diukur dan divalidasi secara kuantitatif pada tahap evaluasi (*evaluation*) guna memastikan tingkat kehandalan serta akurasi prediksi dari model algoritma yang telah diimplementasikan (Hapsari & Pandya Wisesa, 2025).

### 2.4 Analisis Sentimen

Analisis sentimen merupakan salah satu cabang penelitian terapan dari penambangan teks (*text mining*) yang bertujuan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan kecenderungan polaritas opini maupun emosi dari sebuah dokumen tekstual. Secara komputasional, proses analisis ini membedah data teks bahasa alami untuk mengategorikan narasi ke dalam kelas probabilitas polaritas tertentu, yang secara umum terbagi menjadi sentimen positif, negatif, maupun netral (Josen Limbong et al., 2022). Dalam penerapannya, klasifikasi polaritas sentimen dapat dieksekusi secara otomatis dan sistematis dengan mengimplementasikan model pembelajaran mesin (*machine learning*), salah satunya adalah dengan menggunakan algoritma *Naive Bayes* (Ramadhan As'ari et al., 2025).

Dalam konteks observasi karya musik, khususnya pada pemrosesan teks lirik lagu, analisis sentimen diimplementasikan untuk memetakan nuansa emosional serta perasaan yang secara tersurat maupun tersirat terkandung di dalam karya tersebut (Manueke et al., 2023). Hal ini menjadi sangat esensial karena sebuah lirik lagu merupakan medium representasi perasaan manusia,

sehingga pendekatan komputasional diperlukan untuk menguraikan variasi diksi bahasa tersebut menjadi bentuk kelas emosi yang dapat dipahami oleh mesin (Kholida Zia Abidin et al., 2023; Sabda Husada & Yulianto, 2026). Melalui pendekatan pengklasifikasian teks emosional ini, kecenderungan makna naratif yang membentuk lirik lagu dapat diidentifikasi secara kuantitatif dan diinterpretasikan secara objektif (Syafaah & Lestandy, 2022).

## 2.5 Ekstraksi Kata Kunci

Ekstraksi kata kunci merupakan salah satu metode analisis dalam pemrosesan bahasa komputasional yang difungsikan untuk mengidentifikasi dan menemukan kata-kata paling dominan dari sebuah korpus teks. Dalam disiplin ilmu penambangan teks (*text mining*), implementasi ekstraksi kata kunci berperan krusial untuk meringkas dan merepresentasikan *term* yang paling mendasar dalam suatu dokumen tekstual (Odianto et al., 2025). Melalui metode ini, sistem secara otomatis akan memetakan probabilitas kemunculan fitur kata sehingga inti narasi dari teks yang bersifat tidak terstruktur (*unstructured data*) dapat dipahami secara komprehensif tanpa harus mengevaluasi keseluruhan isi dokumen secara manual.

Pada praktiknya, hasil komputasi dari proses ekstraksi kata kunci sering kali direpresentasikan ke dalam bentuk pemodelan grafis untuk mempermudah interpretasi analisis, salah satunya adalah dengan menggunakan teknik visualisasi awan kata (*word cloud*) (Josen Limbong et al., 2022). Pendekatan visualisasi ini bekerja dengan prinsip proporsionalitas pembobotan, di mana kata-kata yang memiliki frekuensi dan bobot tertinggi di dalam keseluruhan korpus teks akan ditampilkan menggunakan dimensi tipografi yang lebih besar dan mencolok. Dalam konteks penelitian lirik lagu, ekstraksi kata dominan tersebut menjadi parameter yang objektif untuk mempertegas identifikasi tema narasi utama, yang kemudian melengkapi hasil klasifikasi polaritas pada analisis sentimen.

## 2.6 Lirik Lagu Sebagai Objek Penelitian

Secara fundamental, lirik lagu merupakan medium ekspresi artistik yang esensial, di mana susunan kata demi kata sering kali dirangkai untuk merangkum emosi terdalam manusia—seperti rasa sepi yang ditenangkan oleh kehadiran suatu sosok—menjadi sebuah karya naratif yang bermakna. Dalam konteks penelitian komputasional, untaian bahasa kiasan dan ungkapan perasaan yang membentuk lirik lagu tersebut telah menjadi salah satu objek observasi utama dalam disiplin ilmu *Natural Language Processing* (NLP) dan penambangan teks (*text mining*) (Manueke et al., 2023). Karakteristik leksikal pada lirik lagu yang sarat akan diksi menjadikannya sebagai representasi data tekstual yang sangat ideal untuk dianalisis guna memetakan polaritas emosi dan melakukan pengklasifikasian sentimen secara objektif (Kholida Zia Abidin et al., 2023; Sabda Husada & Yulianto, 2026).

Lebih lanjut, menjadikan lirik lagu sebagai objek penelitian memberikan peluang yang krusial untuk mengimplementasikan algoritma pembelajaran mesin (*machine learning*) dan pendekatan sains data (*data science*)

(Rosebaugh & Shamir, 2022). Melalui pendekatan teknologi tersebut, peneliti dapat mengekstraksi dan mengenali pola-pola bahasa yang mewakili kondisi emosional tertentu dari pencipta lagu (Ode et al., 2022). Transformasi lirik dari sebuah karya seni sastra menjadi matriks data numerik memungkinkan sistem komputasi untuk membedah, mengklasifikasikan, serta mengevaluasi kedalaman makna teks yang terkandung di dalam karya musik tersebut.

### **2.6.1 Pengertian Lirik Lagu**

Lirik lagu secara konseptual merupakan susunan kata-kata tertulis yang diciptakan secara spesifik untuk diiringi oleh instrumen atau aransemen musik. Dalam perspektif linguistik dan ilmu komunikasi, lirik lagu tidak sekadar berfungsi sebagai pelengkap melodi, melainkan bertindak sebagai medium penyampaian pesan, ekspresi, serta wacana pragmatis dari seorang pencipta karya kepada para pendengarnya (Salsabila Nuzula & Djauhari, 2025). Teks pada sebuah lagu sering kali memuat struktur narasi komprehensif yang merepresentasikan pengalaman personal, realitas sosial, hingga kondisi emosional tertentu yang dikemas melalui penggunaan diksi yang estetik.

Sebagai sebuah dokumen tekstual, lirik lagu memiliki karakteristik leksikal yang unik apabila dikomparasikan dengan teks literatur formal lainnya. Hal ini tidak terlepas dari kebebasan seniman atau musisi independen dalam merangkai kata demi kata untuk menghasilkan rima, irama, dan resonansi emosi yang kuat di setiap baitnya (Rosebaugh & Shamir, 2022). Oleh karena itu, lirik lagu diakui sebagai salah satu bentuk karya sastra kontemporer yang sangat sarat akan makna dan nilai afektif, sehingga menjadikannya objek yang sangat ideal bagi sistem komputasi linguistik untuk meneliti dan mengklasifikasikan ungkapan emosi yang tertuang di dalamnya (Kholida Zia Abidin et al., 2023; Sabda Husada & Yulianto, 2026).

### **2.6.2 Genre R&B**

*Rhythm and Blues* (R&B) merupakan *genre* musik populer yang secara historis memiliki akar perpaduan antara elemen *jazz*, *gospel*, dan *blues*. Dalam perkembangannya, R&B modern telah bertransformasi menjadi salah satu pilar utama dalam industri musik dengan karakteristik komposisi yang berfokus pada ritme vokal dan penyampaian melodi yang halus (*smooth*). Pada konteks linguistik dan struktur narasi, lirik lagu bergenre R&B sangat kental dengan artikulasi perasaan dan wacana emosional yang mendalam (Salsabila Nuzula & Djauhari, 2025). Teks pada karya R&B sering kali membedah kompleksitas dinamika hubungan manusia, mulai dari romansa, kerinduan, hingga wacana permohonan (*plea*), yang dikonstruksi melalui gaya bahasa pragmatis dan diksi ekspresif (Salsabila Nuzula & Djauhari, 2025). Intensitas penyampaian lirik yang sangat emosional ini menjadikan R&B sebagai objek korpus teks yang sangat kaya akan nilai polaritas untuk dieksekusi dalam analisis sentimen.

### 2.6.3 Genre *Neo-Soul*

*Neo-soul* diakui sebagai *sub-genre* atau bentuk evolusi kontemporer dari pergerakan musik *soul* dan R&B yang diklasifikasikan melalui peleburan elemen *soul* klasik dengan instrumen *jazz*, *funk*, dan elemen *hip-hop* alternatif (Hider & Lee, 2023). Berbeda dengan musik populer pada umumnya, karakteristik *neo-soul* cenderung mengedepankan aransemen yang lebih organik, di mana substansi liriknya dirancang untuk selaras dengan eksplorasi akor musik yang kompleks. Secara tekstual, lirik lagu *neo-soul* memiliki kecenderungan gaya bahasa yang jauh lebih puitis, introspektif, dan memiliki tingkat kesadaran (*conscious*) yang tinggi secara batiniah. Lirik pada genre ini sering kali mengeksplorasi spektrum narasi yang menenangkan jiwa, pencarian jati diri, hingga pengamatan sosial, menjadikannya sebuah dokumen sastra dengan kompleksitas teks (*soulful*) yang ideal untuk diurai melalui tahapan ekstraksi kata kunci pada pemrosesan bahasa alami (Hider & Lee, 2023).

### 2.6.4 Karakteristik Lirik *R&B* dan *Neo-Soul*

Karakteristik utama dari teks lirik pada karya musik R&B dan *neo-soul* terletak pada kedalaman eksplorasi emosi serta penggunaan gaya bahasa yang sangat ekspresif dan introspektif (Salsabila Nuzula & Djauhari, 2025). Secara leksikal, perpaduan komposisi dari kedua *genre* ini menghasilkan korpus teks yang tidak sekadar menarasikan sebuah peristiwa secara harfiah, melainkan berfokus pada resonansi perasaan batiniah dan kesadaran (*conscious*) dari sang pencipta karya (Rosebaugh & Shamir, 2022). Teks lirik sering kali dirangkai menggunakan diksi yang puitis guna merangkum pengalaman personal, mulai dari wacana tentang dinamika hubungan, permohonan (*plea*), pencarian jati diri, hingga tema-tema yang mendalam seperti merangkum kehadiran suatu sosok yang mampu menenangkan hati yang sedang kesepian (Hider & Lee, 2023; Salsabila Nuzula & Djauhari, 2025).

Dari perspektif ilmu komputasi, karakteristik naratif yang sangat kompleks (*soulful*) pada lirik R&B dan *neo-soul* menjadikannya sekumpulan data tekstual (*dataset*) yang sangat kaya akan nilai probabilitas bagi sistem pemrosesan bahasa alami. Penguraian lirik-lirik yang sarat akan makna kiasan ini membutuhkan tahapan penambangan teks (*text mining*) yang akurat agar tema-tema dominan di dalamnya dapat dipetakan secara terstruktur. Melalui pendekatan tersebut, esensi afektif dan ungkapan emosional yang menjadi ciri khas lirik pada *genre* ini dapat dikonversi menjadi representasi kelas polaritas sentimen yang terukur, objektif, dan dapat dipahami secara logis oleh mesin (Sabda Husada & Yulianto, 2026).

## 2.7 Pre-processing

*Pre-processing* merupakan tahapan krusial dalam *text mining* yang berfungsi untuk membersihkan dan menstandarisasi data tekstual mentah (*raw data*) sebelum diproses lebih lanjut oleh model komputasi. Data teks yang

diekstraksi, seperti lirik lagu, umumnya masih bersifat tidak terstruktur dan mengandung banyak gangguan (*noise*), seperti variasi huruf, tanda baca, serta kata-kata yang tidak memiliki nilai analitis (Manueke et al., 2023). Oleh karena itu, tahapan pra-pemrosesan diimplementasikan untuk mengubah teks bahasa alami tersebut menjadi format yang lebih terstruktur dan seragam, sehingga algoritma pembelajaran mesin (*machine learning*) dapat mengekstraksi fitur kata dengan tingkat akurasi yang lebih optimal (Sabda Husada & Yulianto, 2026).

### **2.7.1 Case Folding**

*Case folding* adalah proses penyeragaman seluruh karakter huruf pada korpus teks menjadi bentuk huruf kecil (*lowercase*). Proses ini bertujuan untuk menghindari inkonsistensi pembacaan matriks oleh sistem, karena secara komputasional, karakter huruf kapital dan huruf kecil sering kali direpresentasikan dengan nilai biner yang berbeda. Melalui implementasi *case folding*, kata yang letaknya di awal kalimat maupun di tengah kalimat akan dievaluasi sebagai satu entitas string yang sama dalam tahapan perhitungan bobot selanjutnya (Ramadhan As'ari et al., 2025).

### **2.7.2 Cleaning**

*Cleaning* atau pembersihan merupakan tahapan filtrasi untuk menghapus elemen-elemen teks yang tidak relevan dan tidak memberikan kontribusi makna terhadap proses analisis emosi. Pada tahapan ini, seluruh karakter non-alfabetik seperti angka, tanda baca, simbol (*special characters*), hingga elemen seperti tautan (*URL*) akan dieliminasi dari dokumen (Manueke et al., 2023). Pembersihan ini dilakukan untuk memastikan bahwa sistem komputasi hanya akan memproses susunan kata murni yang menyusun struktur lirik lagu tersebut.

### **2.7.3 Tokenizing**

*Tokenizing* merupakan tahapan pemecahan atau pemotongan serangkaian teks kalimat utuh menjadi unit-unit kata tunggal yang berdiri sendiri, yang dikenal sebagai token. Proses ini umumnya memisahkan *string* teks berdasarkan spasi atau karakter pemisah lainnya. Pemecahan kalimat menjadi token tunggal merupakan langkah esensial karena perhitungan probabilitas dan klasifikasi sentimen pada algoritma akan dieksekusi berdasarkan frekuensi kemunculan dari masing-masing unit token tersebut (Tri Widodo & Setyawan, 2025).

### **2.7.4 Stopword Removal**

*Stopword removal* adalah tahapan penyaringan untuk mengeliminasi kata-kata umum yang sering muncul di dalam teks namun tidak memiliki bobot makna spesifik untuk keperluan klasifikasi polaritas. Contoh kata yang masuk dalam daftar *stopword* meliputi kata hubung, kata ganti, atau kata depan, seperti "dan", "di", "ke", "yang", atau "dari". Dengan menghapus *stopword*, dimensi ruang data akan berkurang secara signifikan

sehingga proses komputasi mesin menjadi lebih efisien dan difokuskan secara khusus pada kata kunci yang mendefinisikan emosi (Sabda Husada & Yulianto, 2026).

### 2.7.5 Stemming

*Stemming* adalah proses analisis morfologis yang bertujuan untuk mengubah token kata yang memiliki imbuhan (prefiks, sufiks, infiks, maupun konfiks) kembali menjadi bentuk kata dasarnya (*root word*). Sebagai representasi, sistem akan mentransformasi variasi turunan kata menjadi entitas kata dasar tunggal. Tahapan ini beroperasi secara esensial dalam pemrosesan teks bahasa Indonesia guna mengurangi redundansi kamus data dan memastikan bahwa variasi leksikal yang berasal dari akar makna yang sama akan dikalkulasi sebagai satu fitur probabilitas yang ekuivalen (Ramadhan As'ari et al., 2025).

### 2.8 Term Frequency-Inverse Document Frequency (TF-IDF)

*Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan salah satu algoritma pembobotan kata yang paling fundamental dalam disiplin *text mining* dan pemrosesan bahasa alami. Metode ini digunakan untuk mengevaluasi dan mengkalkulasi seberapa penting sebuah kata (*term*) terhadap suatu dokumen tertentu di dalam sebuah kumpulan korpus teks (Virginia & Irsyad, 2025). Dalam proses komputasi, teks lirik lagu yang telah melewati tahapan pra-pemrosesan (*pre-processing*) akan diekstraksi nilai bobot numeriknya menggunakan algoritma ini, sehingga mesin dapat mengidentifikasi relevansi dari setiap kata sebelum diklasifikasikan ke dalam kelas sentimen (Syafaah & Lestandy, 2022).

Pendekatan komputasi pada metode ini mengintegrasikan dua parameter matematis utama. Parameter pertama adalah *Term Frequency* (TF), yang berfungsi untuk mengukur frekuensi kemunculan sebuah kata spesifik di dalam satu dokumen tunggal. Semakin sering kata tersebut muncul di dalam lirik, semakin tinggi nilai probabilitas lokalnya. Parameter kedua adalah *Inverse Document Frequency* (IDF), yang bertindak sebagai fungsi normalisasi untuk mengurangi atau memberi penalti pada bobot kata-kata umum yang muncul di hampir seluruh dokumen dalam korpus (Virginia & Irsyad, 2025). Melalui perpaduan parameter ini, TF-IDF mampu menyoroti kata kunci unik yang menjadi representasi utama dari dokumen tersebut. Secara matematis, kalkulasi nilai bobot akhir (*weight*) dari setiap kata diperoleh dengan mengalikan nilai frekuensi kemunculan (TF) dan nilai keunikan kata (IDF). Persamaan matematis pembobotan TF-IDF dijabarkan sebagai berikut:

$$W_{t,d} = TF(t,d) \times IDF(t)$$

Untuk memperoleh nilai probabilitas balikan dokumen (IDF), rumusan logaritmik yang digunakan adalah:

$$IDF(t) = \log\left(\frac{N}{df(t)}\right)$$

Keterangan

$W_{t,d}$  : Bobot akhir representasi kata (*term*)  $t$  dalam dokumen  $d$

$TF(t, d)$  : Bobot akhir representasi kata (*term*)  $t$  dalam dokumen  $d$

$N$  : Total keseluruhan jumlah dokumen di dalam korpus teks

$df(t)$  : Jumlah total dokumen yang mengandung kata (*term*)  $t$

## 2.9 Algoritma *Naïve Bayes*

Algoritma *Naive Bayes* merupakan salah satu metode klasifikasi probabilitas terawasi (*supervised learning*) yang berakar dari konsep statistika Teorema Bayes (Riyandona et al., 2025). Dalam bidang penambangan teks (*text mining*), algoritma ini sangat populer dan secara luas diimplementasikan untuk mengeksekusi proses analisis sentimen (Tri Widodo & Setyawan, 2025). Penggunaan *Naive Bayes* dinilai sangat relevan untuk mengklasifikasikan lirik lagu karena algoritma ini memiliki efisiensi komputasi yang tinggi dan kehandalan dalam memproses kumpulan data dengan dimensi atribut yang sangat besar (Ramadlan As'ari et al., 2025).

Pendekatan klasifikasi pada algoritma ini disebut "*naive*" atau lugu karena didasarkan pada asumsi penyederhanaan yang menganggap bahwa setiap atribut atau fitur kata di dalam sebuah dokumen bersifat independen yang kuat antarkondisinya (Ode et al., 2022). Asumsi independensi ini berarti bahwa probabilitas kemunculan sebuah kata di dalam teks dianggap tidak memiliki keterikatan struktural atau dampak langsung terhadap kemunculan kata lainnya. Meskipun kondisi kebebasan atribut tersebut jarang terjadi secara mutlak dalam tata bahasa alami, klasifikasi *Naive Bayes* terbukti mampu mengalkulasi dan memprediksi kelas polaritas emosi dengan tingkat akurasi yang tinggi (Josen Limbong et al., 2022).

Secara matematis, klasifikasi teks dieksekusi dengan mengalkulasi nilai probabilitas tertinggi dari suatu kelas target (misalnya sentimen positif, negatif, atau netral) berdasarkan keberadaan fitur-fitur kata pendukungnya. Persamaan probabilitas pada metode *Naive Bayes* dirumuskan melalui persamaan berikut:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

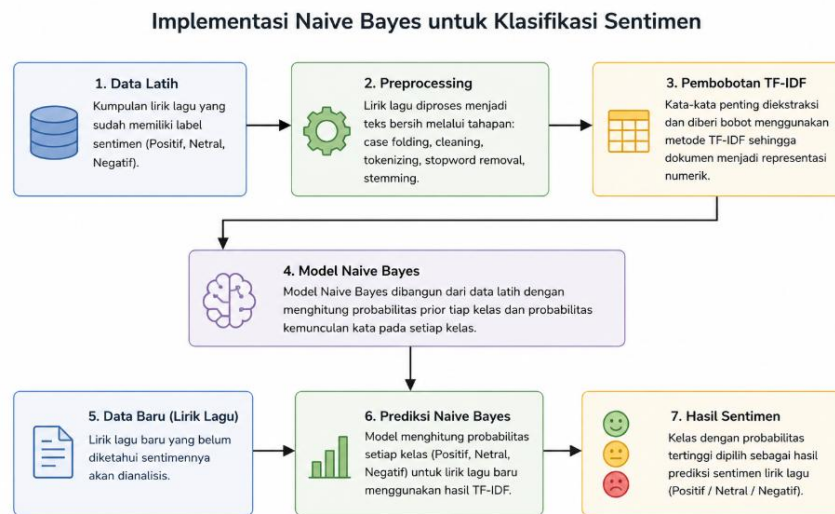
Keterangan:

$P(C|X)$  : Peluang bersyarat dari kelas sentimen  $C$  apabila diketahui dokumen  $X$  (*posterior probability*).

$P(X|C)$  : Peluang kemunculan fitur dokumen  $X$  pada kelas sentimen  $C$  (*likelihood*).

$P(C)$  : Peluang kemunculan kelas sentimen  $C$  secara global (*prior probability*).

$P(X)$  : Peluang kemunculan dokumen  $X$  secara global (*evidence*).



**Gambar 2. 1 Alur Implementasi Metode *Multinomial Naive Bayes***

## 2.10 Evaluasi Model

Tahapan evaluasi model merupakan fase krusial dalam siklus komputasi yang difungsikan untuk mengukur tingkat kinerja, kehandalan, serta validitas dari algoritma pembelajaran mesin (*machine learning*) yang telah diimplementasikan (Hapsari & Pandya Wisesa, 2025). Pada penelitian berbasis klasifikasi teks, pengujian ini bertujuan untuk memvalidasi seberapa akurat sistem dalam memprediksi kelas polaritas emosi jika dikomparasikan dengan label aktual atau data sesungguhnya (Riyandona et al., 2025). Hasil dari evaluasi ini akan merepresentasikan kelayakan sistem komputasi dalam mengekstraksi dan mengklasifikasikan lirik lagu secara objektif.

### 2.10.1 *Confusion Matrix*

*Confusion matrix* atau matriks kebingungan adalah alat pengujian performa berupa representasi tabel yang memvisualisasikan matriks perbandingan antara hasil prediksi algoritma komputasi dengan data target atau label kelas aktual (Riyandona et al., 2025). Tabel ini memberikan rincian komprehensif mengenai tingkat keberhasilan maupun kesalahan sistem dalam mengklasifikasikan teks. Matriks ini mengekstraksi empat parameter dasar pengujian, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Keempat parameter inilah yang nantinya dikalkulasikan untuk menghasilkan metrik evaluasi kuantitatif secara lebih spesifik.

### 2.10.2 *Accuracy*

*Accuracy* atau akurasi adalah metrik pengukuran yang merepresentasikan persentase ketepatan atau kecocokan keseluruhan hasil prediksi model algoritma terhadap total keseluruhan data teks yang diuji

(Hapsari & Pandya Wisesa, 2025). Metrik ini menjadi indikator umum yang paling sering diandalkan untuk melihat seberapa baik sistem membedakan setiap kelas sentimen secara global. Secara matematis, tingkat akurasi dirumuskan melalui persamaan berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Keterangan:

**TP** (*True Positive*) : Jumlah data polaritas positif yang berhasil diprediksi dengan benar sebagai positif oleh sistem.

**TN** (*True Negative*) : Jumlah data polaritas negatif yang berhasil diprediksi dengan benar sebagai negatif oleh sistem.

**FP** (*False Positive*) : Jumlah data polaritas negatif yang salah diprediksi sebagai positif oleh sistem.

**FN** (*False Negative*): Jumlah data polaritas positif yang salah diprediksi sebagai negatif oleh sistem.

### 2.10.3 Precision

*Precision* atau presisi merupakan metrik spesifik yang mengukur tingkat ketepatan prediksi positif yang dihasilkan oleh sistem (Riyandona et al., 2025). Metrik ini menunjukkan rasio antara jumlah data yang diprediksi benar positif oleh model dengan total seluruh data yang diprediksi sebagai positif (baik yang benar maupun yang salah prediksi). Persamaan matematis untuk kalkulasi *precision* adalah sebagai berikut:

$$Precision = \frac{TP}{TP + FP}$$

### 2.10.4 Recall

*Recall* atau tingkat kelengkapan adalah metrik komputasional yang mengalkulasi rasio jumlah prediksi benar positif dibandingkan dengan total keseluruhan data aktual yang bernilai positif di dalam korpus uji (Hapsari & Pandya Wisesa, 2025). Metrik ini sangat esensial untuk mengetahui sejauh mana algoritma mampu mengenali dan memanggil kembali (*recall*) seluruh informasi yang relevan di dalam sebuah basis data tekstual. Rumusan matematis *recall* dijabarkan sebagai berikut:

$$Recall = \frac{TP}{TP + FN}$$

### 2.10.5 F1-Score

*F1-Score* merupakan metrik evaluasi yang mempresentasikan nilai rata-rata harmonik dari *precision* dan *recall* (Hapsari & Pandya Wisesa, 2025). Penggunaan metrik ini sangat relevan apabila terdapat distribusi kelas sampel yang tidak seimbang (*imbalanced data*) di dalam *dataset*.

Melalui kalkulasi *F1-Score*, peneliti dapat memperoleh representasi tunggal yang menyeimbangkan *trade-off* antara nilai presisi dan tingkat kelengkapan pemanggilan data. Persamaan kalkulasi *F1-Score* dirumuskan melalui persamaan berikut:

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Keterangan:

**Precision** : Tingkat rasio ketepatan prediksi positif dari sistem.

**Recall** : Tingkat kelengkapan sistem dalam memprediksi data positif secara aktual.

## 2.11 Studi Literatur

Studi literatur berfungsi untuk mengkaji penelitian terdahulu guna membangun landasan teoretis, mengevaluasi metodologi komputasi, serta mengidentifikasi celah penelitian (*research gap*) dan nilai kebaruan (*novelty*). Pada penelitian ini, rujukan literatur dikerucutkan pada sepuluh jurnal utama yang berfokus pada implementasi *Natural Language Processing* (NLP), penambangan teks (*text mining*), algoritma *Naive Bayes*, serta pemrosesan teks pada lirik lagu. Komparasi dari kesepuluh literatur terdahulu tersebut dijabarkan secara sistematis pada Tabel 2.1 berikut.

Tabel 2. 1 Studi Literatur

| No | Peneliti & Tahun               | Judul Penelitian                                                                            | Metode/Algoritma                               | Hasil Penelitian                                                                                                                           | Kontribusi terhadap Riset                                                                                                    |
|----|--------------------------------|---------------------------------------------------------------------------------------------|------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| 1  | Manueke et al. (2023)          | Analisis Sentimen Lirik Lagu Populer Dengan Pemanfaatan Teknik Natural Language Processing. | Natural Language Processing (NLP), Text Mining | Implementasi NLP terbukti efektif dalam mentransformasi teks lirik lagu yang ambigu menjadi data yang dapat diekstraksi nilai sentimennya. | Menjadi rujukan landasan teori bahwa NLP merupakan metode yang paling tepat untuk memproses struktur bahasa pada lirik lagu. |
| 2  | Sabda Husada & Yulianto (2026) | Analisis Emosi dalam Lirik Lagu menggunakan Natural Language Processing.                    | Natural Language Processing (NLP)              | Sistem komputasi berhasil mengenali dan mengklasifikasikan ungkapan emosi spesifik dari berbagai                                           | Memberikan landasan observasi mengenai pemetaan kelas polaritas emosi pada teks sastra kontemporer seperti lirik lagu.       |

|   |                                  |                                                                                              |                                     |                                                                                                                                     |                                                                                                                                     |
|---|----------------------------------|----------------------------------------------------------------------------------------------|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
|   |                                  |                                                                                              |                                     | objek teks karya musik.                                                                                                             |                                                                                                                                     |
| 3 | Ode et al. (2022)                | Klasifikasi Lirik Lagu Bertema Lingkungan dengan Metode Naive Bayes.                         | Naive Bayes Classifier              | Algoritma Naive Bayes menghasilkan tingkat akurasi yang tinggi saat diterapkan pada korpus dokumen lirik bertema khusus.            | Memvalidasi kehandalan algoritma Naive Bayes dalam memproses sekumpulan data teks lirik yang memiliki dimensi atribut sangat besar. |
| 4 | Kholida Zia Abidin et al. (2023) | Systematic Literature Review terhadap Klasifikasi Emosi pada Lirik Lagu.                     | Systematic Literature Review (SLR)  | Pemetaan berbagai variasi algoritma pembelajaran mesin terbukti sangat krusial dalam menganalisis ungkapan afektif lirik lagu.      | Menjustifikasi urgensi penelitian analisis emosi pada karya musik dan membantu menentukan posisi kebaruan (novelty) penelitian ini. |
| 5 | Riyandona et al. (2025)          | Implementasi Model Analisis Sentimen Terhadap Grup Musik BTS Menggunakan Metode Naive Bayes. | Naive Bayes, Analisis Sentimen      | Model probabilitas bersyarat pada Naive Bayes terbukti sangat efisien dalam memproses sentimen opini publik terkait industri musik. | Memperkuat argumen pemilihan algoritma probabilitas karena ketepatan metrik evaluasinya pada objek bertema musik.                   |
| 6 | Josen Limbong et al. (2022)      | Analisis Klasifikasi Sentimen Ulasan Berbasis Word Cloud dengan Metode Naive Bayes.          | Naive Bayes, Visualisasi Word Cloud | Penggabungan hasil klasifikasi dengan awan kata (word cloud) sangat efektif memetakan probabilitas kata dominan dari teks.          | Menjadi referensi utama untuk perancangan fitur ekstraksi topik dan visualisasi kata kunci (term) pada antarmuka sistem.            |
| 7 | Odhianto et al. (2025)           | Optimalisasi Latent                                                                          | Text Mining,                        | Ekstraksi topik secara akurat                                                                                                       | Memberikan referensi mengenai                                                                                                       |

|    |                                    |                                                                                       |                                      |                                                                                                                                     |                                                                                                                                    |
|----|------------------------------------|---------------------------------------------------------------------------------------|--------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
|    |                                    | Dirichlet Allocation untuk Ekstraksi Topik Utama dalam Teks Dongeng.                  | Ekstraksi Topik                      | berhasil menemukan tema naratif tersembunyi pada dokumen sastra yang tidak terstruktur.                                             | pentingnya ekstraksi kata kunci untuk merangkum tema utama pada karya berbasis teks puitis/sastra.                                 |
| 8  | Virginia & Irsyad (2025)           | Rekomendasi Lagu Berdasarkan Jenis Mood Menggunakan TF-IDF.                           | Pembobotan TF-IDF, Cosine Similarity | Parameter bobot TF-IDF mampu memetakan frekuensi kata yang paling relevan dengan suasana hati (mood) dari sebuah lagu.              | Menjadi landasan matematis penerapan algoritma pembobotan TF-IDF untuk menyeleksi fitur kata pada lirik R&B.                       |
| 9  | Salsabila Nuzula & Djauhari (2025) | Discourse of Apology and Plea in R&B Lyrics: A Pragmatic Analysis.                    | Analisis Pragmatis, Linguistik       | Secara leksikal, lirik lagu R&B terbukti memiliki kedalaman wacana emosional tinggi, sarat akan diksi permohonan dan introspeksi.   | Memberikan landasan karakteristik linguistik dan justifikasi kuat mengapa lirik R&B dan neo-soul dipilih sebagai objek penelitian. |
| 10 | Ramadhan As'ari et al. (2025)      | Analisis Sentimen Ulasan Pengguna Aplikasi Spotify Menggunakan Algoritma Naive Bayes. | Naive Bayes, Pre-processing Teks     | Tahapan pra-pemrosesan yang dipadukan dengan Naive Bayes menghasilkan evaluasi matriks kebingungan (confusion matrix) yang optimal. | Sebagai acuan tahapan sistematis pra-pemrosesan teks (pre-processing) untuk membersihkan gangguan (noise) pada data linguistik.    |

## BAB III METODOLOGI PENELITIAN

### 3.1 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental yang berfokus pada penerapan komputasi berbasis teks atau *text mining*. Pendekatan kuantitatif dipilih karena analisis variasi emosi dan klasifikasi polaritas dalam penelitian ini diukur serta dieksekusi menggunakan pemodelan statistik berbasis angka, yang dihasilkan dari perhitungan matematis pembobotan tekstual lirik lagu. Metode utama yang diterapkan dalam penelitian ini adalah penambangan teks (*text mining*), sebuah disiplin turunan dari *data mining* yang secara spesifik mengekstraksi pola informasi berharga, pengetahuan baru, atau karakteristik tersembunyi dari sekumpulan data tekstual yang bersifat tidak terstruktur (*unstructured data*).

#### 3.1.1 Kerangka Penelitian

Guna memastikan proses penambangan teks berjalan secara sistematis, penelitian ini mengadopsi kerangka kerja *Knowledge Discovery in Databases* (KDD) (Syafaah & Lestandy, 2022). Kerangka kerja *Knowledge Discovery in Databases* (KDD) dipilih karena memiliki tahapan terstruktur yang sangat adaptif dalam mengolah data dari bentuk mentah hingga menjadi pengetahuan (*knowledge*) yang siap diinterpretasikan. Secara umum, alur kronologis penelitian komputasi lirik lagu ini berjalan melalui tahapan-tahapan makro sebagai berikut:

1. **Fase Seleksi Data (*Data Selection*):** Melakukan pengumpulan data lirik lagu R&B dan *neo-soul* dari platform digital, kemudian menyaring dokumen tersebut berdasarkan parameter inklusi untuk membentuk korpus *dataset* yang valid dan memenuhi standar minimal kuantitas analisis.
2. **Fase Pra-Pemrosesan (*Pre-processing*):** Mentransformasikan data lirik mentah yang sarat akan gangguan (*noise*) ke dalam bentuk teks yang bersih dan seragam melalui teknik sekuensial *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*.
3. **Fase Transformasi (*Transformation*):** Mengonversi kata dasar hasil pembersihan menjadi matriks koordinat vektor numerik menggunakan metode pembobotan frekuensi dan keunikan kata melalui algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF).
4. **Fase Penambangan Teks (*Text Mining*):** Mengimplementasikan algoritma klasifikasi berbasis probabilitas *Naive Bayes* pada matriks numerik untuk menentukan label polaritas sentimen pada dokumen lirik, serta melakukan identifikasi topik berdasarkan kemunculan kata-kata dominan yang merepresentasikan tema utama pada korpus lirik.
5. **Fase Evaluasi (*Evaluation*):** Mengukur performa model klasifikasi menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengetahui tingkat keberhasilan dan kehandalan algoritma dalam melakukan klasifikasi sentimen.

6. **Fase Perancangan Antarmuka Sistem:** Merancang struktur menu dan tampilan layar aplikasi sebagai media interaksi pengguna dalam melakukan pengolahan data lirik, analisis sentimen, serta visualisasi hasil ekstraksi topik.

### 3.1.2 Arsitektur Sistem

Sistem klasifikasi sentimen dan identifikasi topik dominan ini dibangun menggunakan arsitektur gabungan (*hybrid*) berbasis web. Pengembangan antarmuka pengguna (*Graphical User Interface*) dan pengelolaan struktur basis data relasional (MySQL) diimplementasikan menggunakan bahasa pemrograman PHP yang bertindak sebagai gerbang interaksi utama sistem. Di sisi lain, keseluruhan komputasi kecerdasan buatan—yang mencakup tahapan pra-pemrosesan teks, pembobotan TF-IDF, ekstraksi fitur dominan, hingga pelokasian model Multinomial Naive Bayes—dieksekusi secara terpusat oleh mesin *Machine Learning* menggunakan bahasa pemrograman Python. Kedua lingkungan pengembangan perangkat lunak ini diintegrasikan secara dinamis melalui fungsi jembatan *shell execution* (*shell\_exec*), sehingga sistem mampu membaca pembaruan korpus lirik, melatih ulang model probabilitas, serta memprediksi sentimen data secara *real-time* langsung melalui peramban (*browser*).

## 3.2 Dataset dan Sumber Data Lirik Lagu

Penelitian ini menggunakan *dataset* sekunder berupa sekumpulan data teks lirik lagu kontemporer yang secara spesifik diklasifikasikan ke dalam genre R&B dan *neo-soul*. Proses pengumpulan data lirik utama dilakukan melalui kombinasi pengumpulan manual dan pemanfaatan sumber data lirik digital dari platform pangkalan data lirik musik terbesar, yakni Genius.com. Pemilihan Genius.com sebagai sumber data didasarkan pada tingkat akurasi liriknya yang divalidasi langsung oleh komunitas dan musisi, sehingga sangat meminimalisir kesalahan penulisan teks (*typo*) pada korpus.

Di samping penarikan data secara otomatis, pengumpulan data juga dilakukan secara manual untuk mengakomodasi karya-karya dari ekosistem musik independen lokal yang relevan dengan genre tersebut. Total keseluruhan data yang berhasil dihimpun dan digunakan dalam penelitian ini adalah sebanyak **207 baris dokumen lirik lagu**. Angka tersebut telah dipastikan memenuhi ambang batas minimal kelayakan pemrosesan *text mining* sesuai dengan standar pedoman akademik.

*Dataset* yang telah dikurasi tersebut kemudian disimpan ke dalam format *Comma Separated Values* (.csv) agar kompatibel dan mudah diimpor melalui fitur *browse* pada antarmuka *Graphical User Interface* (GUI) sistem yang dirancang. Setiap baris data pada korpus ini memuat tiga atribut utama yang menjadi parameter analisis, yaitu nama artis, judul lagu, dan bait lirik mentah (*raw text*) yang memuat keseluruhan isi lagu.

Tabel 3.1 di bawah ini menampilkan cuplikan representatif dari *dataset* lirik lagu yang digunakan dalam sistem, mencakup rentang karya

dari musisi arus utama internasional hingga salah satu sampel karya independen lokal yang merepresentasikan karakteristik narasi *neo-soul* secara personal:

**Tabel 3. 1 Sampel Dataset Lirik Lagu**

| ID | Artis                | Judul Lagu         | Lirik Lagu (Data Teks Mentah)                                           |
|----|----------------------|--------------------|-------------------------------------------------------------------------|
| 1  | Mac Ayres            | Easy               | It is easy to fall in love with you but hard to stay...                 |
| 2  | xdika                | Teduh              | Langkahmu lirik namun, menyibak gelap tanpa ragu...                     |
| 3  | Daniel Caesar        | Get You            | Through drought and famine natural disasters my baby has been around... |
| 4  | Glenn Fredly         | Sedih Tak Berujung | Saat menjelang hari-hari bahagiamu aku memilih diam dalam sepi...       |
| 5  | Maliq & D'Essentials | Untitled           | Tidakkah cukup yang engkau lihat pertemanan ini sendiri...              |

### 3.3 Seleksi Data (*Data Selection*)

Tahap seleksi data (*data selection*) merupakan langkah prosedural pertama dan fundamental untuk membangun fondasi informasi sebelum masuk ke tahapan komputasi bahasa alami. Pada fase ini, dilakukan pengumpulan sekaligus penyaringan (*filtering*) terhadap sekumpulan data teks mentah (*raw data*) yang bersumber dari platform Genius.com, yang dipadukan dengan hasil kurasi lirik secara manual. Penggunaan sumber data yang komprehensif ini bertujuan untuk memperoleh variasi leksikal yang kaya, sekaligus mengakomodasi karya-karya dari musisi arus utama (*mainstream*) maupun karya dari ekosistem musik independen yang bergerak secara spesifik di jalur R&B dan *neo-soul*.

Mengingat volume data tekstual di platform digital sangat masif dan bersifat heterogen, proses seleksi diterapkan menggunakan parameter inklusi dan eksklusi yang ketat. Proses utama difokuskan pada isolasi genre, di mana sistem hanya mengekstrak lirik yang secara eksplisit dikategorikan ke dalam *metadata* genre R&B dan *neo-soul* kontemporer. Selanjutnya, dilakukan penyaringan anomali data untuk mengeliminasi lagu-lagu instrumental (tanpa lirik), versi pertunjukan langsung (*live version*), maupun versi paduan (*remix*) yang berpotensi menghasilkan duplikasi kalimat atau pergeseran topik. Teks lirik yang terlalu pendek atau tidak memiliki struktur bait yang utuh juga dibuang dari korpus data untuk mencegah terjadinya bias (*noise*) pada saat algoritma melakukan pembobotan kata di tahap selanjutnya.

Selain penyaringan kualitas tekstual, tahapan ini juga mencakup proses penghapusan data ganda (*data deduplication*) untuk memastikan bahwa setiap judul lagu yang masuk ke dalam *dataset* merupakan entitas analisis yang unik. Proses seleksi ini dikurasi secara ketat guna memverifikasi bahwa *dataset* akhir telah memenuhi dan melampaui ambang batas minimal kelayakan pemrosesan

*text mining* sesuai dengan standar kewajiban akademik, yakni mencapai hasil akhir sebanyak **207 baris** dokumen representatif.

Korpus data lirik pilihan yang telah terverifikasi tersebut kemudian diseragamkan struktur atributnya—yang mencakup nama artis, judul lagu, dan teks lirik—untuk selanjutnya disimpan secara terpusat ke dalam format berkas *Comma Separated Values* (.csv). Format ini dipastikan kompatibel untuk dieksekusi pada tahapan pra-pemrosesan (*pre-processing*).

### 3.4 Pra-Pemrosesan (*Pre-Processing*)

Tahap Pra-pemrosesan data merupakan tahapan krusial dalam komputasi *text mining* yang bertujuan untuk membersihkan dan menormalisasi *dataset* teks mentah sebelum masuk ke tahap ekstraksi fitur. Karena dokumen lirik lagu umumnya mengandung struktur teks yang tidak baku, karakter spesial, serta banyak kosakata yang tidak memiliki bobot sentimen, tahap pra-pemrosesan mutlak diperlukan untuk mencegah penurunan tingkat akurasi pada algoritma Multinomial Naive Bayes. Dalam penelitian ini, proses pra-pemrosesan diimplementasikan menggunakan bahasa pemrograman Python dengan memanfaatkan pustaka pemrosesan bahasa alami *Natural Language Toolkit* (NLTK) dan Sastrawi melalui lima tahapan sekuensial berikut:

#### 3.4.1 *Case Folding*

Tahap *case folding* adalah proses penyeragaman seluruh karakter huruf di dalam dokumen teks menjadi huruf kecil (*lowercase*). Pada tahap ini, mesin mengonversi seluruh huruf kapital yang terdapat pada awal kalimat, nama, maupun penekanan lirik menjadi huruf kecil. Hal ini bertujuan agar algoritma komputasi menganggap kata yang sama dengan penulisan kapitalisasi yang berbeda (contoh: "*Love*", "*LOVE*", dan "*love*") sebagai satu entitas kata yang identik.

**Tabel 3. 2 *Case Folding***

| Sebelum <i>Case Folding</i>                                                  | Sesudah <i>Case Folding</i>                                                  |
|------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| Saat menjelang hari-hari bahagiamu, aku memilih diam dalam sepi... Oh, yeah! | saat menjelang hari-hari bahagiamu, aku memilih diam dalam sepi... oh, yeah! |

#### 3.4.2 *Pembersihan (Cleaning)*

Tahap *cleaning* merupakan proses pembersihan dokumen dari karakter-karakter yang tidak memiliki nilai analitik. Menggunakan metode *Regular Expression* (Regex) pada Python, sistem secara otomatis menghapus seluruh angka, tanda baca, simbol, serta karakter non-alfabet lainnya. Pada konteks *dataset* lirik lagu, tahap ini secara spesifik sangat berguna untuk menghilangkan tanda kurung siku atau kurung lengkung yang biasanya memuat penanda struktur

lagu seperti [*Chorus*], [*Verse*], maupun (*Intro*) yang sering terbawa saat proses penarikan data ( *scraping* ).

**Tabel 3. 3 Cleaning**

| Sebelum <i>Cleaning</i>                                                            | Sesudah <i>Cleaning</i>                                                      |
|------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| saat menjelang hari-hari<br>bahagiamu, aku memilih diam<br>dalam sepi... oh, yeah! | saat menjelang hari hari<br>bahagiamu aku memilih diam<br>dalam sepi oh yeah |

### 3.4.3 Tokenizing

Setelah teks lirik bersih dari karakter non-alfabet dan seragam dalam format huruf kecil, langkah selanjutnya adalah *tokenizing*. Tahap ini memecah untaian kalimat panjang di dalam lirik menjadi potongan-potongan kata tunggal atau *token* yang berdiri sendiri. Pemisahan ini umumnya menggunakan spasi sebagai karakter pembatas (*delimiter*), sehingga mesin dapat menghitung frekuensi dan menganalisis probabilitas masing-masing kata secara individual pada tahapan pemodelan.

**Tabel 3. 4 Tokenizing**

| Sebelum <i>Tokenizing</i>                                                    | Sesudah <i>Tokenizing</i>                                                                                            |
|------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|
| saat menjelang hari hari<br>bahagiamu aku memilih diam<br>dalam sepi oh yeah | ["saat", "menjelang", "hari",<br>"hari", "bahagiamu", "aku",<br>"memilih", "diam", "dalam",<br>"sepi", "oh", "yeah"] |

### 3.4.4 Penghapusan Kata Umum (*Stopword Removal*)

*Stopword removal* adalah tahap penyaringan untuk membuang kata-kata umum yang sering muncul namun tidak membawa informasi atau makna sentimen yang signifikan (seperti kata hubung, kata ganti, dan kata depan). Sistem ini menerapkan penyaringan berlapis menggunakan kamus bawaan dari NLTK untuk bahasa Inggris dan Sastrawi untuk bahasa Indonesia.

Sebagai bentuk optimasi pada penelitian ini, ditambahkan pula sebuah kamus *custom stopwords* yang dirancang khusus untuk mengeliminasi *noise* leksikal pada genre musik R&B dan *neo-soul*. Kata-kata dominan yang bersifat gumaman vokal, *ad-libs*, atau ekspresi musikal yang tidak memiliki muatan sentimen (seperti *yeah*, *ooh*, *mmm*, *uh*, *ah*) secara otomatis dihapus dari korpus. Pembuangan *noise* spesifik ini terbukti krusial untuk mencegah algoritma mengalami kebingungan (*overlapping features*) dalam membedakan kelas sentimen positif dan negatif.

**Tabel 3. 5 Stopword Removal**

| Sebelum <i>Stopword Removal</i>                                                                             | Sesudah <i>Stopword Removal</i>                                       |
|-------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| ["saat", "menjelang", "hari", "hari", "bahagiamu", "aku", "memilih", "diam", "dalam", "sepi", "oh", "yeah"] | ["menjelang", "hari", "hari", "bahagiamu", "memilih", "diam", "sepi"] |

### 3.4.5 Stemming / Lemmatization

Tahap terakhir dari proses pra-pemrosesan adalah *stemming*, yaitu proses pemotongan atau transformasi *token* kata yang memiliki imbuhan (afiksasi) kembali ke bentuk kata dasarnya (*root word*). Proses ini dilakukan menggunakan algoritma aturan morfologi dari pustaka Sastrawi. Sebagai contoh, kata-kata yang memiliki variasi imbuhan akan diseragamkan menjadi satu bentuk dasar, sehingga mengurangi keragaman kosakata (*dimensionality reduction*) tanpa menghilangkan maknanya. Hasil akhir dari tahap *stemming* ini adalah *dataset* lirik yang telah bersih, padat, dan siap untuk dikonversi menjadi matriks angka melalui pembobotan TF-IDF.

**Tabel 3. 6 Stemming**

| Sebelum <i>Stemming</i>                                               | Sesudah <i>Stemming</i>                                        |
|-----------------------------------------------------------------------|----------------------------------------------------------------|
| ["menjelang", "hari", "hari", "bahagiamu", "memilih", "diam", "sepi"] | ["jelang", "hari", "hari", "bahagia", "pilih", "diam", "sepi"] |

### 3.5 Ekstraksi Fitur Menggunakan TF-IDF

Pada tahap ini, korpus teks lirik yang telah melewati seluruh proses pembersihan akan dikonversi menjadi representasi numerik menggunakan algoritma pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF). Proses ekstraksi fitur ini sangat krusial karena mesin *Machine Learning* tidak dapat memproses data dalam bentuk karakter alfabet murni, melainkan harus diubah ke dalam bentuk matriks vektor angka (Josen Limbong et al., 2022).

Secara matematis, TF-IDF mengevaluasi seberapa relevan dan penting sebuah kata di dalam satu lirik lagu relatif terhadap seluruh kumpulan lagu (*dataset*). *Term Frequency* (TF) bertugas menghitung jumlah frekuensi kemunculan sebuah kata di dalam satu lirik. Sementara itu, *Inverse Document Frequency* (IDF) bertindak sebagai penyeimbang dengan cara memberikan "penalti" atau penurunan bobot pada kata-kata umum yang muncul hampir di semua lagu, sehingga kata-kata spesifik yang unik akan mendapatkan bobot numerik yang lebih tinggi (Josen Limbong et al., 2022).

Lebih lanjut, sebagai bentuk ekstraksi informasi lanjutan dari penelitian ini, perhitungan frekuensi kata (*Term Frequency*) tidak hanya digunakan sebagai masukan untuk model klasifikasi, melainkan juga dimanfaatkan untuk melakukan identifikasi topik dominan. Berdasarkan

hasil pemeringkatan matriks TF-IDF, kosakata yang memiliki kemunculan tertinggi (seperti kata *love*, *baby*, *feel*, *time*, *life*) akan diekstraksi menjadi representasi tema utama dari korpus lirik. Pemunculan kata-kata dominan yang berkaitan dengan hubungan interpersonal ini membuktikan bahwa ekstraksi topik pada lirik lagu bergenre R&B dan *neo-soul* secara valid merefleksikan narasi romansa dan gejolak emosional (Salsabila Nuzula & Djauhari, 2025).

### 3.6 Klasifikasi Sentimen Menggunakan Multinomial Naive Bayes

Setelah dokumen teks berhasil direpresentasikan dalam bentuk matriks bobot numerik melalui TF-IDF, tahap komputasi inti selanjutnya adalah melakukan prediksi polaritas sentimen menggunakan algoritma probabilitas *Multinomial Naive Bayes* (MNB). Pemilihan varian *multinomial* dalam arsitektur sistem ini didasarkan pada kecocokannya secara matematis untuk menangani fitur data diskrit berupa perhitungan frekuensi, sehingga sangat ideal untuk disandingkan dengan matriks keluaran dari TF-IDF (Riyandona et al., 2025).

Secara konseptual, algoritma *Naive Bayes* bekerja berlandaskan pada Teorema Bayes dengan mengedepankan asumsi "naif" (*naive*), yakni menganggap bahwa kemunculan setiap fitur kata di dalam lirik lagu saling independen dan tidak memiliki keterkaitan dengan keberadaan kata lainnya (Riyandona et al., 2025). Sistem akan mengalkulasi probabilitas kemunculan matriks kata dari lirik yang diuji terhadap data latih (*training data*) yang telah diberi label (positif, netral, atau negatif) sebelumnya.

Dokumen lirik lagu baru yang dimasukkan ke dalam sistem kemudian akan diklasifikasikan ke dalam kelas sentimen yang mengantongi nilai probabilitas *posterior* (probabilitas akhir) tertinggi. Penggunaan *Multinomial Naive Bayes* diterapkan secara spesifik pada penelitian ini mengingat komputasinya yang sangat efisien dalam memori, memiliki waktu komputasi yang cepat, serta terbukti memiliki performa akurasi yang tangguh dalam menyelesaikan masalah klasifikasi pada kumpulan data teks (*text mining*) berdimensi tinggi, khususnya dalam analisis emosi lirik lagu (Sabda Husada & Yulianto, 2026).

### 3.7 Rancangan Basis Data

Rancangan basis data merupakan tahapan perancangan arsitektur penyimpanan data yang digunakan untuk menampung seluruh informasi sistem. Sistem klasifikasi sentimen ini menggunakan sistem manajemen basis data relasional MySQL. Mengingat sistem ini berfokus pada pemrosesan teks (*text mining*), arsitektur basis data dirancang secara efisien menggunakan dua entitas tabel utama, yaitu tabel admin untuk autentikasi keamanan dan tabel *dataset* untuk menyimpan korpus lirik beserta hasil transformasinya.

#### 3.7.1 Logical Record Structure (LRS)

Representasi struktur data secara logis tanpa ketergantungan pada mekanisme penyimpanan fisik didefinisikan melalui konsep *Logical Record Structure* (LRS). Model ini difungsikan untuk memvisualisasikan skema relasional antar-entitas, di mana integritas referensial data dijaga

melalui mekanisme pengaitan antara kunci utama (*primary key*) dan kunci tamu (*foreign key*). Pemahaman mengenai struktur ini penting untuk memastikan bahwa setiap elemen data dapat diakses dan dikelola secara efisien dalam sistem basis data relasional.

Dalam konteks sistem yang dibangun, implementasi LRS dirancang untuk menjamin kontinuitas alur informasi yang terintegrasi di dalam basis data tugasakhir\_andika, yang difokuskan pada relasi antara entitas *tbl\_admin* dan *tbl\_dataset*. Konektivitas antar-tabel dibangun menggunakan atribut identitas unik (*id\_admin*) yang berperan sebagai jembatan penghubung antara data autentikasi pengguna dengan pengelolaan korpus lirik lagu beserta seluruh riwayat transformasi teksnya. Konfigurasi lengkap mengenai struktur logika relasional tersebut disajikan pada Gambar 3.1.



**Gambar 3.1 Logical Record Structure**

### 3.7.2 Class Diagram

Untuk memodelkan struktur statis dari sistem basis data yang dibangun, digunakan *Class Diagram* yang merupakan bagian dari standar *Unified Modeling Language* (UML). Diagram ini memvisualisasikan skema sistem dengan memetakan kelas-kelas entitas, atribut yang dimilikinya, serta relasi antar kelas tersebut. Dalam konteks penelitian analisis sentimen lirik ini, *Class Diagram* berfungsi menggambarkan alur penyimpanan data, mulai dari tahap autentikasi hingga pencatatan hasil klasifikasi akhir. Struktur rinci dari diagram tersebut direpresentasikan pada Gambar 3.2.



**Gambar 3. 2 Class Diagram**

Berdasarkan rancangan visual di atas, sistem basis data dikonstruksi atas dua entitas utama yang saling terintegrasi. Deskripsi fungsional dari masing-masing entitas adalah sebagai berikut:

a. **tbl\_admin**

Entitas ini berperan sebagai repositori utama untuk menampung data kredensial pengguna yang memiliki hak akses penuh ke dalam sistem. Atribut kunci yang disimpan meliputi *id\_admin* bertipe *integer* sebagai penanda unik, serta *username* dan *password* bertipe *varchar* untuk proses autentikasi keamanan. Kelas ini memegang relasi *one-to-many* (1:M) terhadap entitas dataset, di mana satu admin memiliki wewenang untuk mengelola dan memproses banyak dokumen lirik lagu.

b. **tbl\_dataset**

Tabel ini difungsikan sebagai repositori utama atau basis pengetahuan (*knowledge base*) yang memuat rekapitulasi korpus data lirik beserta seluruh jejak rekam pra-pemrosesan teksnya. Informasi yang disimpan mencakup *id* sebagai identitas unik, *id\_admin* sebagai atribut penghubung, serta *metadata* musik berupa *artist*, *title*, dan *genre*. Lebih lanjut, entitas ini mendokumentasikan setiap tahapan transformasi *text mining*, mulai dari lirik mentah (*lyrics*), luaran *case\_folding*, *cleaning*, *tokenizing*, *stopword\_removed*, hingga *stemmed* dan *clean\_lyrics*. Selain itu, kronologi penyimpanan dicatat dalam atribut *upload\_date*, sedangkan parameter target dan prediksi dari algoritma klasifikasi direkam secara terstruktur pada atribut *sentiment\_label* dan *sentiment\_result*.

### 3.7.3 Spesifikasi Basis Data

Spesifikasi basis data menjabarkan rincian tipe data, panjang karakter, dan fungsi dari masing-masing atribut (kolom) yang terdapat di

dalam tabel basis data. Berikut adalah spesifikasi dari kedua tabel yang digunakan dalam sistem:

a. Tabel *tbl\_admin*

Tabel *tbl\_admin* berfungsi sebagai repositori sentral untuk autentikasi dalam arsitektur sistem ini. Tabel ini didesain khusus untuk menampung seluruh himpunan data pengguna yang memiliki hak akses (*privilege*) sebagai pengelola sistem. Identitas dan karakteristik teknis dari tabel ini adalah sebagai berikut:

Nama Database : tugasakhir\_andika

Nama Table : *tbl\_admin*

Primary Key : *id\_admin*

Isi : Data Autentikasi Pengelola Sistem

Untuk memberikan gambaran yang lebih komprehensif mengenai struktur penyimpanan, rincian atribut kolom yang mencakup nama *field*, tipe data, kapasitas penyimpanan (lebar), serta deskripsi fungsional masing-masing atribut disajikan dalam Tabel 3.7.

**Tabel 3. 7 Struktur Tabel *tbl\_admin***

| Nama Field      | Tipe Data | Panjang | Keterangan                      |
|-----------------|-----------|---------|---------------------------------|
| <i>id_admin</i> | int       | 11      | ID unik pengelola sistem (PK)   |
| <i>username</i> | varchar   | 50      | Nama pengguna untuk autentikasi |
| <i>password</i> | varchar   | 255     | Kata sandi pengguna             |

b. Tabel *tbl\_dataset*

Tabel *tbl\_dataset* dikonstruksi sebagai entitas referensi utama atau wadah komputasi teks (*text mining*) dalam kerangka penelitian ini. Fungsi fundamental dari tabel ini adalah untuk mengarsipkan data lirik lagu mentah beserta rekam jejak hasil pra-pemrosesan dan label sentimennya. Adapun spesifikasi teknis yang diterapkan pada tabel ini adalah sebagai berikut:

Nama Database : tugasakhir\_andika

Nama Table : *tbl\_dataset*

Primary Key : *id*

Foreign Key : *id\_admin*

Isi : Data Lirik Lagu dan Hasil *Text Mining*

Guna memberikan pemahaman mendalam mengenai atribut yang digunakan dalam proses komputasi, rincian lengkap terkait nama *field*, tipe data, serta deskripsi fungsional setiap kolom disajikan dalam Tabel 3.8.

**Tabel 3. 8 Struktur Tabel *tbl\_dataset***

| Nama Field | Tipe Data | Panjang | Keterangan                 |
|------------|-----------|---------|----------------------------|
| <i>id</i>  | int       | 11      | ID unik dokumen lirik (PK) |

|                  |           |     |                                       |
|------------------|-----------|-----|---------------------------------------|
| id_admin         | int       | 11  | ID pengelola sistem (FK)              |
| artist           | varchar   | 255 | Nama artis pengubah lagu              |
| title            | varchar   | 255 | Judul lagu                            |
| genre            | varchar   | 100 | Kategori genre musik lagu tersebut    |
| lyrics           | longtext  |     | Data teks mentah dari lirik lagu      |
| case_folding     | text      |     | Lirik hasil konversi huruf kecil      |
| cleaning         | text      |     | Lirik bebas angka dan tanda baca      |
| tokenizing       | text      |     | Teks lirik yang dipecah menjadi token |
| stopword_removed | text      |     | Teks bebas noise dan kata hubung      |
| stemmed          | text      |     | Lirik dalam bentuk kata dasar         |
| upload_date      | timestamp |     | Waktu data diunggah ke sistem         |
| clean_lyrics     | text      |     | Lirik bersih untuk proses mesin       |
| sentiment_label  | varchar   | 20  | Target kelas prediksi aktual          |
| sentiment_result | varchar   | 20  | Hasil tebakan dari Naive Bayes        |

### 3.8 Rancangan Pengujian

Tahapan pengujian dirancang secara sistematis guna mengevaluasi efektivitas sistem klasifikasi sentimen lirik lagu, baik dari aspek fungsionalitas perangkat lunak maupun akurasi pemodelan bahasa alami terhadap data korpus lirik. Proses evaluasi ini dijalankan melalui implementasi aplikasi berbasis web yang mengintegrasikan basis data tugasakhir\_andika ke dalam satu ekosistem penyimpanan. Integrasi tersebut difungsikan untuk memfasilitasi eksekusi seluruh rangkaian proses, mulai dari *preprocessing*, ekstraksi fitur pembobotan TF-IDF, hingga penerapan algoritma *Multinomial Naive Bayes*.

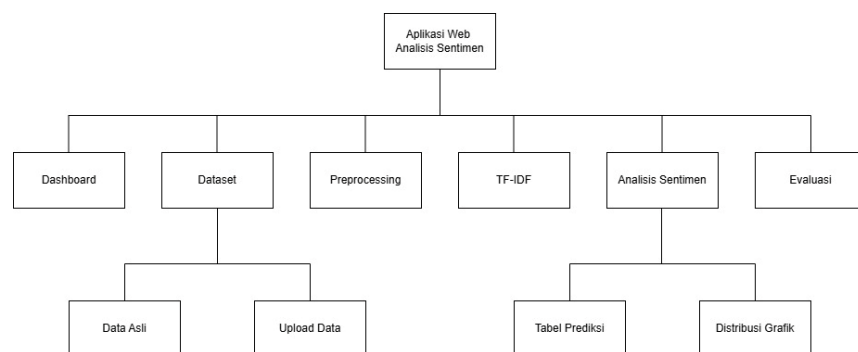
Evaluasi kinerja sistem dibagi menjadi dua dimensi utama. Pertama, penilaian fungsionalitas sistem (*Blackbox Testing*) dilakukan tanpa melibatkan struktur kode internal aplikasi, melainkan berfokus pada pengujian interaksi antarmuka pengguna seperti kelancaran proses unggah *dataset* CSV, eksekusi pembersihan teks, dan navigasi menu. Penggunaan metode pengujian ini bertujuan untuk mengukur tingkat kehandalan dan stabilitas sistem sebelum dioperasikan secara penuh.

Kedua, validasi performa model komputasi dilakukan dengan mengomparasikan hasil tebakan kelas sentimen (positif, netral, negatif) yang dikeluarkan oleh mesin terhadap label sentimen aktual (*ground truth*) dari dokumen lirik tersebut. Indikator keberhasilan pada tahap ini diukur secara kuantitatif melalui implementasi *Confusion Matrix* menggunakan parameter *Accuracy*, *Precision*, *Recall*, serta *F1-Score*. Melalui pendekatan pengujian yang komprehensif ini, diharapkan dapat diperoleh gambaran utuh mengenai ketepatan sistem komputasi dalam mengenali polaritas emosi pada teks lirik musik R&B dan *neo-soul*.

### 3.9 Rancangan Menu

Rancangan menu merupakan representasi visual dari arsitektur informasi dan struktur navigasi antarmuka pada sistem yang akan dibangun. Perancangan ini dipetakan secara konseptual dalam bentuk diagram pohon hierarki (*sitemap*) yang berfungsi untuk memvisualisasikan seluruh fungsionalitas aplikasi secara logis dan terstruktur. Melalui rancangan navigasi ini, alur interaksi pengguna dan tahapan perpindahan antar halaman dapat dipahami dengan jelas. Hal tersebut menjadi esensial untuk memberikan panduan penggunaan yang intuitif, mulai dari tahap inisiasi data hingga visualisasi hasil akhir pemrosesan penambangan teks (*text mining*).

Dalam pengembangannya, struktur menu pada sistem analisis lirik lagu ini dirancang secara sekuensial guna mengakomodasi alur komputasi eksperimen secara utuh. Arsitektur sistem ini mencakup enam modul utama, yaitu *Dashboard*, *Dataset*, *Pre-processing*, *TF-IDF*, *Analisis Sentimen*, dan *Evaluasi Model*. Khusus pada modul *Dataset*, struktur dikembangkan menjadi sistem hierarki bertingkat (*dropdown*) yang memuat dua sub-menu fungsional, yaitu *Upload Dataset* dan *Data Lirik*. Pembagian dan pengelompokan menu tersebut memastikan bahwa setiap proses transformasi data tekstual memiliki ruang eksekusi yang spesifik dan terorganisasi dengan baik. Bentuk visual dari rancangan arsitektur menu sistem tersebut dapat dilihat pada Gambar 3.3 berikut:



**Gambar 3.3 Rancangan Menu**

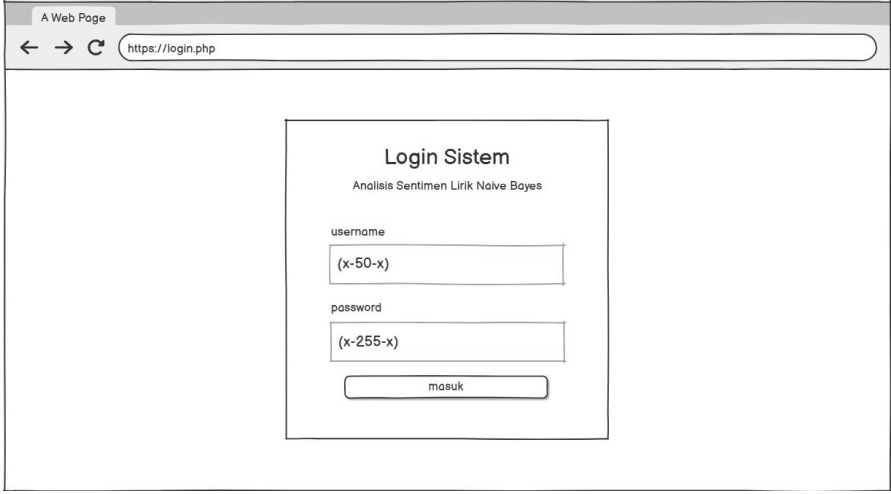
### 3.10 Rancangan Layar

Rancangan layar (*User Interface Design*) merupakan representasi visual dari antarmuka sistem web yang akan dibangun. Tujuan dari perancangan ini

adalah untuk memberikan gambaran arsitektur informasi dan tata letak (*layout*) elemen-elemen interaktif, sehingga pengoperasian fungsionalitas sistem oleh pengguna dapat berjalan secara intuitif dan efisien. Berikut adalah rancangan antarmuka untuk masing-masing halaman pada sistem klasifikasi sentimen lirik lagu *R&B* dan *Neo-Soul*:

### 3.10.1 Rancangan Layar Login

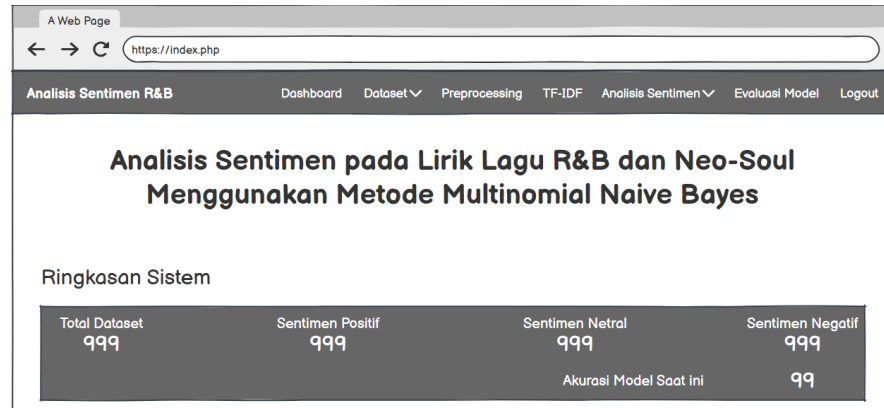
Halaman *login* merupakan gerbang autentikasi pertama bagi administrator untuk dapat mengakses sistem. Pada antarmuka ini, terdapat formulir yang mewajibkan pengguna untuk memasukkan kredensial berupa *username* dan *password* yang valid sebelum mendapatkan hak akses ke halaman utama.

The image shows a web browser window with the address bar displaying 'https://login.php'. The main content area contains a login form titled 'Login Sistem' with the subtitle 'Analisis Sentimen Lirik Naive Bayes'. The form has two input fields: 'username' with a placeholder '(x-50-x)' and 'password' with a placeholder '(x-255-x)'. Below these fields is a button labeled 'masuk'.

Gambar 3. 4 Rancangan Layar Login

### 3.10.2 Rancangan Layar Dashboard

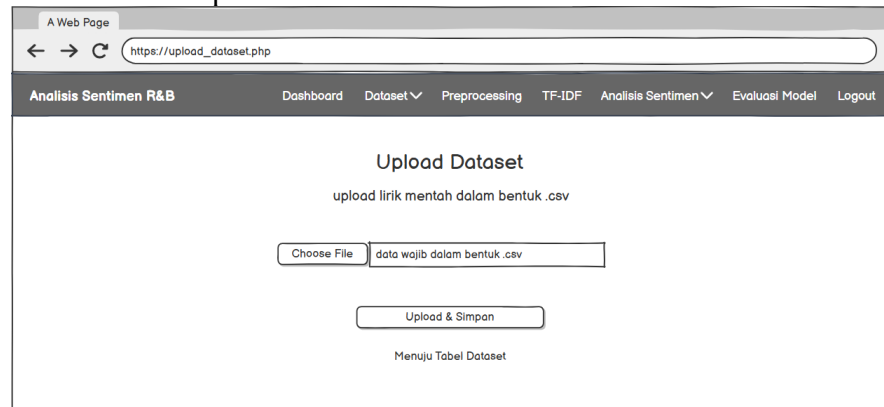
Halaman *dashboard* dirancang sebagai panel informasi utama yang pertama kali dilihat oleh admin setelah berhasil melewati proses autentikasi. Halaman ini berfungsi untuk menyajikan ringkasan singkat berupa indikator angka atau kartu informasi (*information cards*) terkait keseluruhan distribusi data yang ada di dalam basis data sistem.



**Gambar 3. 5 Rancangan Layar Dashboard**

### 3.10.3 Rancangan Layar Upload CSV

Halaman ini disediakan secara khusus sebagai modul masukan (*input*) data korpus. Admin dapat mencari, memilih, dan mengunggah berkas kumpulan lirik lagu dalam format CSV (*Comma Separated Values*) melalui tombol *Select File* sebelum data tersebut direkam secara permanen.



**Gambar 3. 6 Rancangan Layar Upload CSV**

### 3.10.4 Rancangan Layar Data Lirik Asli (*Raw Dataset*)

Halaman *Raw Dataset* berfungsi untuk menampilkan daftar dokumen lirik R&B dan Neo-Soul secara utuh yang belum mengalami proses modifikasi komputasional. Tabel pada halaman ini menyajikan detail informasi yang meliputi ID, judul lagu, nama artis, cuplikan lirik mentah (*snippet*), serta label sentimen manual yang menjadi acuan pengujian.

A Web Page

← → ↻ https://dataset.php

Analisis Sentimen R&B Dashboard Dataset ▼ Preprocessing TF-IDF Analisis Sentimen ▼ Evaluasi Model Logout

**Data Lirik Asli**  
Menampilkan Data Lirik Mentah Sebelum Preprocessing

| ID    | Judul Lagu | Artis     | Lirik Mentah | Label Manual |
|-------|------------|-----------|--------------|--------------|
| 9(11) | (x-255-x)  | (x-255-x) | (x-n-x)      | (x-20-x)     |
|       |            |           |              |              |
|       |            |           |              |              |
|       |            |           |              |              |
|       |            |           |              |              |

**Gambar 3. 7 Rancangan Layar Dataset Raw**

### 3.10.5 Rancangan Layar *Pre-Processing*

Halaman ini menyajikan tabel komparasi transparan mengenai rekam jejak pembersihan teks (*text preprocessing*). Melalui tabel ini, admin dapat memantau transformasi lirik lagu secara bertahap dari bentuk mentah hingga tereduksi melalui proses *case folding*, *cleaning*, *tokenizing*, dan *stopword removed*.

A Web Page

← → ↻ https://preprocessing.php

Analisis Sentimen R&B Dashboard Dataset ▼ Preprocessing TF-IDF Analisis Sentimen ▼ Evaluasi Model Logout

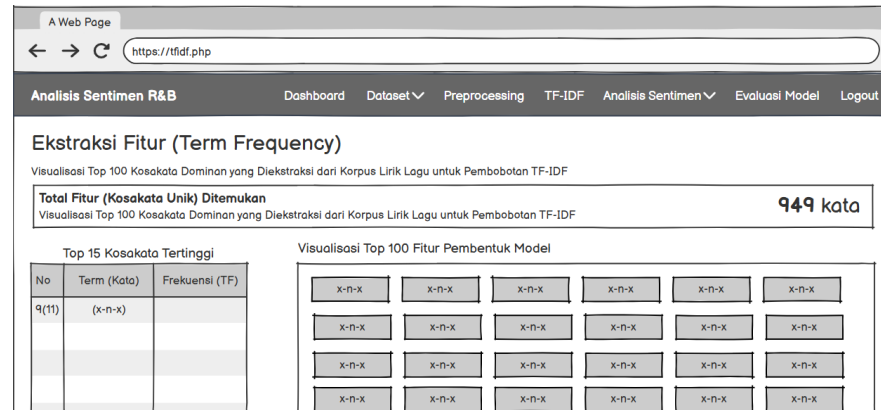
**Hasil Teks Preprocessing**

| No    | Info Lagu | Lirik Mentah | Case Folding | Cleaning | Tokenizing | Stopword Removal | Stemming |
|-------|-----------|--------------|--------------|----------|------------|------------------|----------|
| 9(11) | (x-255-x) | (x-n-x)      | (x-n-x)      | (x-n-x)  | (x-n-x)    | (x-n-x)          | (x-n-x)  |
|       |           |              |              |          |            |                  |          |
|       |           |              |              |          |            |                  |          |
|       |           |              |              |          |            |                  |          |
|       |           |              |              |          |            |                  |          |

**Gambar 3. 8 Rancangan Layar Pre-processing**

### 3.10.6 Rancangan Layar Ekstraksi Fitur (TF-IDF)

Halaman ekstraksi fitur memperlihatkan hasil kalkulasi pembobotan kata menggunakan algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF). Tampilan ini memuat informasi total kosakata unik yang berhasil diekstrak, tabel daftar kata dengan frekuensi kemunculan tertinggi, serta representasi visual dari fitur kata yang dominan dalam membentuk model.



**Gambar 3. 9 Rancangan Layar Ekstraksi Fitur**

### 3.10.7 Rancangan Layar Tabel Hasil Klasifikasi

Halaman hasil klasifikasi mendeskripsikan luaran utama komputasi dari algoritma *Multinomial Naive Bayes*. Tabel yang disajikan memuat hasil prediksi polaritas sentimen terhadap dokumen teks lirik untuk kemudian dianalisis ketepatan prediksinya.

**Tabel Hasil Klasifikasi Naive Bayes**

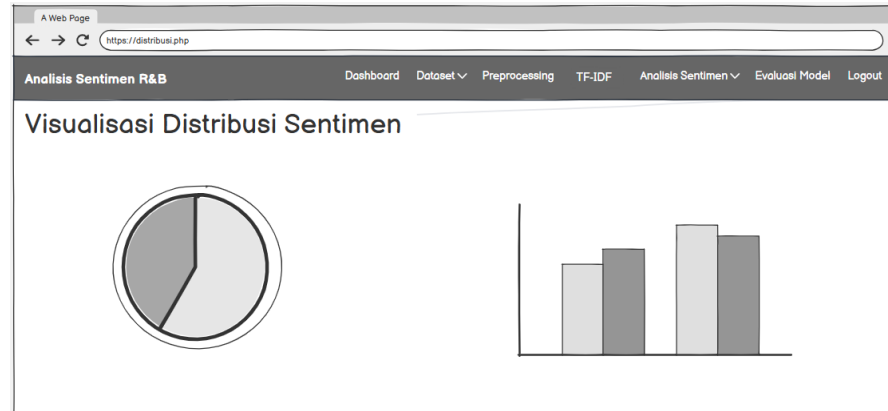
Perbandingan antara kenyataan label asli dan hasil tebakan sistem

| No    | Judul Lagu | Artis     | Kenyataan Asli | Tebakan AI | Status Analisis |
|-------|------------|-----------|----------------|------------|-----------------|
| 9(11) | (x-255-x)  | (x-255-x) | (x-20-x)       | (x-20-x)   | (x-20-x)        |
|       |            |           |                |            |                 |
|       |            |           |                |            |                 |
|       |            |           |                |            |                 |
|       |            |           |                |            |                 |
|       |            |           |                |            |                 |
|       |            |           |                |            |                 |
|       |            |           |                |            |                 |
|       |            |           |                |            |                 |
|       |            |           |                |            |                 |
|       |            |           |                |            |                 |

**Gambar 3. 10 Rancangan Layar Tabel Hasil Klasifikasi**

### 3.10.8 Rancangan Layar Visualisasi Distribusi Sentimen

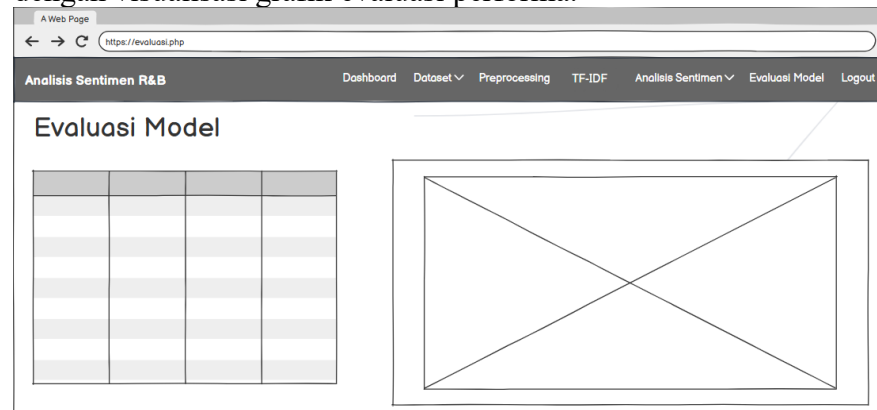
Sebagai pelengkap analisis prediktif, halaman ini merangkum sebaran hasil pelabelan sentimen ke dalam bentuk grafik visual. Terdapat representasi berupa diagram lingkaran (*pie chart*) dan diagram batang (*bar chart*) untuk memudahkan pembacaan komparasi antara kuantitas sentimen pada korpus.



**Gambar 3. 11 Rancangan Layar Visualisasi Distribusi Sentimen**

### 3.10.9 Rancangan Layar Evaluasi Model

Halaman terakhir ini merupakan panel evaluasi yang merangkum tingkat kehandalan algoritma klasifikasi. Antarmuka ini menampilkan parameter matriks kebingungan (*Confusion Matrix*) serta indikator metrik pengukuran kinerja model, yang didampingi dengan visualisasi grafik evaluasi performa.



**Gambar 3. 12 Rancangan Layar Evaluasi Model**

## BAB IV HASIL DAN PEMBAHASAN

### 4.1 Lingkungan Percobaan

Lingkungan percobaan merupakan penjabaran mengenai spesifikasi perangkat keras (*hardware*) dan perangkat lunak (*software*) yang digunakan sebagai infrastruktur komputasi selama proses pengembangan sistem. Lingkungan ini mencakup alat yang digunakan untuk penulisan kode program, pengelolaan basis data, eksekusi algoritma *Machine Learning*, hingga antarmuka berbasis web.

#### 4.1.1 Spesifikasi Hardware

Implementasi sistem analisis sentimen ini dilakukan menggunakan perangkat keras berupa satu unit laptop dengan spesifikasi sebagai berikut:

- |                 |                           |
|-----------------|---------------------------|
| a. Processor    | : Intel Core i5-13420H    |
| b. RAM          | : 16GB                    |
| c. Storage      | : 512GB                   |
| d. Graphic Card | : NVIDIA GeForce RTX 2050 |

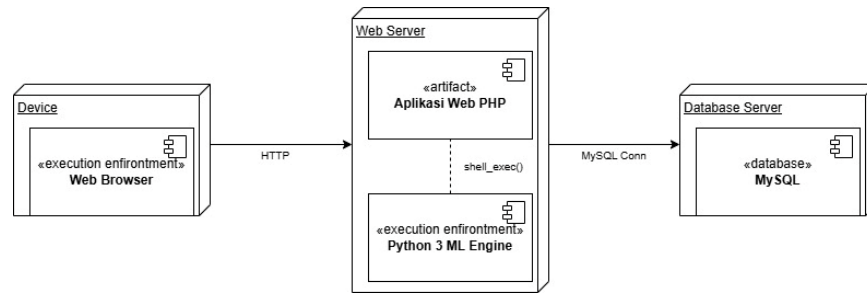
#### 4.1.2 Spesifikasi Software

Perangkat lunak yang digunakan dalam membangun dan menguji sistem ini adalah sebagai berikut:

- |                       |                                             |
|-----------------------|---------------------------------------------|
| a. IDE                | : Visual Studio Code                        |
| b. Browser            | : Microsoft Edge                            |
| c. Bahasa Pemrograman | : Python dan PHP                            |
| d. DBMS               | : MySQL                                     |
| e. OS                 | : Windows 11                                |
| f. Lainnya            | : XAMPP, Microsoft Excel, Draw.io, Balsamiq |

#### 4.1.3 Deployment Diagram

*Deployment diagram* berfungsi untuk memvisualisasikan arsitektur fisik penyebaran komponen perangkat lunak ke dalam simpul (*node*) perangkat keras. Pada sistem ini, arsitektur dibangun menggunakan model *client-server* yang beroperasi pada lingkungan lokal (*localhost*). *Browser* bertindak sebagai *client* yang mengirimkan permintaan (*request*), sementara Apache *Web Server* memproses logika PHP dan menjembatani eksekusi (*shell execution*) *script* Python untuk pemrosesan teks. Seluruh arsip data lirik dan hasil prediksi kemudian dikelola secara terpusat oleh server basis data MySQL.



**Gambar 4. 1 Deployment Diagram**

## 4.2 Implementasi Sistem

Implementasi sistem merupakan tahapan di mana rancangan konsep teoritis dan alur logika diwujudkan ke dalam bentuk operasional komputasi secara nyata. Pada tahapan ini, fokus pembahasan tidak ditekankan pada struktur kode pemrograman secara harfiah, melainkan pada penjabaran proses transformasi data (*Input-Process-Output*). Seluruh proses diurutkan mulai dari penyajian dataset mentah, pembersihan leksikal, ekstraksi matriks pembobotan, hingga hasil akhir prediksi algoritma dan penyimpanannya ke dalam basis data.

### 4.2.1 Implementasi Dataset

Implementasi dataset merupakan tahapan krusial pertama di mana korpus data lirik lagu berbasis teks diinisialisasi ke dalam ekosistem komputasi. Pada fase ini, sistem menerima masukan (input) berupa berkas berformat *Comma Separated Values* (.csv) yang memuat total keseluruhan 207 baris dokumen lirik bergenre R&B dan Neo-Soul.

Berkas dataset mentah tersebut memiliki struktur matriks dua dimensi yang telah dikurasi sebelumnya, mencakup beberapa atribut utama yang berfungsi sebagai parameter analisis awal. Adapun rincian atribut data tersebut adalah sebagai berikut:

- ID** : Merupakan atribut penanda identitas unik (nomor urut) untuk membedakan setiap dokumen lirik di dalam sistem.
- Artis** : Merupakan metadata yang memuat nama musisi, penyanyi, atau band pengubah lagu tersebut.
- Judul Lagu** : Merupakan metadata yang merepresentasikan identitas judul dari karya musik yang sedang dianalisis.
- Lirik Mentah** : Merupakan atribut inti yang memuat untaian teks murni dari keseluruhan bait lagu. Pada tahap ini, teks masih sepenuhnya bersifat tidak terstruktur (*unstructured data*), memiliki variasi kapitalisasi huruf, serta mengandung berbagai gangguan (*noise*) leksikal seperti tanda baca dan simbol.
- Label Manual** : Merupakan parameter target kelas polaritas aktual (Positif, Netral, atau Negatif) yang telah ditetapkan sebelumnya (*ground truth*). Atribut ini akan digunakan sebagai nilai acuan untuk memvalidasi ketepatan hasil prediksi algoritma komputasi pada tahap evaluasi model.

Setelah berkas dataset .csv diunggah melalui menu *Upload* pada antarmuka sistem web, mesin (*server*) akan membaca (*parsing*) keseluruhan struktur matriks tersebut dan menampilkannya sebagai basis data siap proses. Tabel 4.1 berikut mengilustrasikan sampel implementasi dari dataset lirik lagu mentah yang telah berhasil dimuat oleh sistem sebelum memasuki tahapan *pre-processing*:

**Tabel 4. 1 Sampel Implementasi Dataset Mentah**

| ID | Artis                | Judul Lagu         | Lirik Mentah (Raw Data)                                                      | Label Manual |
|----|----------------------|--------------------|------------------------------------------------------------------------------|--------------|
| 1  | Mac Ayres            | Easy               | It is easy to fall in love with you but hard to stay...                      | Positif      |
| 2  | xdika                | Teduh              | Langkahmu lirik namun, menyibak gelap tanpa ragu...                          | Positif      |
| 3  | Daniel Caesar        | Get You            | Through drought and famine natural disasters my baby...                      | Positif      |
| 4  | Glenn Fredly         | Sedih Tak Berujung | Saat menjelang hari hari bahagiamu, aku memilih diam dalam sepi... Oh, yeah! | Negatif      |
| 5  | Maliq & D'Essentials | Untitled           | Tidakkah cukup yang engkau lihat pertemanan ini sendiri...                   | Netral       |

#### 4.2.2 Implementasi Preprocessing

Setelah dataset mentah berhasil dimuat ke dalam sistem, tahapan selanjutnya adalah mengeksekusi proses *preprocessing* (pra-pemrosesan) pada atribut lirik lagu. Implementasi tahapan ini sangat krusial mengingat data lirik yang dikumpulkan masih berupa *unstructured data* yang mengandung banyak gangguan (*noise*) leksikal seperti variasi huruf kapital, tanda baca, simbol, imbuhan kata, serta kosakata umum yang tidak merepresentasikan polaritas sentimen.

Sistem mengimplementasikan lima tahapan *preprocessing* secara sekuensial menggunakan bahasa pemrograman Python, dengan memanfaatkan dukungan pustaka pemrosesan bahasa alami *Natural Language Toolkit* (NLTK) dan Sastrawi. Alur transformasi teks ini dieksekusi langkah demi langkah pada setiap baris dokumen lirik di dalam basis data untuk menghasilkan korpus kata yang bersih, padat, dan seragam.

Adapun lima tahapan yang diimplementasikan pada sistem meliputi:

##### a. *Case Folding*

Tahapan ini merupakan proses penyeragaman seluruh karakter huruf yang terdapat pada dokumen teks lirik lagu menjadi bentuk huruf

kecil (*lowercase*). Dalam pemrosesan bahasa alami, mesin komputasi bersifat sangat peka terhadap format huruf (*case-sensitive*), yang berarti huruf kapital dan huruf kecil dianggap sebagai nilai biner yang berbeda. Oleh karena itu, konversi ini wajib dilakukan untuk memastikan konsistensi pembacaan data, sehingga algoritma tidak mengalami kebingungan saat menghitung frekuensi kata yang sama namun memiliki gaya penulisan kapitalisasi yang berbeda di awal maupun di tengah kalimat.

*b. Cleaning*

Pada tahap pembersihan data ini, sistem secara sistematis menghapus seluruh karakter non-alfabetik dari dalam teks lirik yang tidak memiliki relevansi maupun nilai komputasi. Proses ini dieksekusi untuk mengeliminasi angka, karakter spesial, tautan, hingga berbagai bentuk tanda baca. Pembersihan karakter ini sangat esensial karena elemen-elemen tersebut sering kali hanya berfungsi sebagai pemanis tulisan atau *metadata* tambahan lagu (seperti tanda kurung pada lirik), namun tidak membawa muatan polaritas sentimen apapun bagi model algoritma dan justru berpotensi menjadi gangguan (*noise*).

*c. Tokenizing*

Setelah dokumen lirik dipastikan bersih dari karakter non-alfabetik, proses dilanjutkan dengan pemecahan untaian kalimat panjang menjadi potongan-potongan unit kata tunggal yang disebut sebagai *token*. Sistem mengeksekusi pemecahan ini dengan mengidentifikasi karakter spasi sebagai pemisah utamanya (*delimiter*). Tahapan ini bertindak sebagai jembatan yang sangat krusial sebelum memasuki fase pembobotan, karena algoritma TF-IDF dan *Naive Bayes* hanya dapat menganalisis dan mengkalkulasi probabilitas jika struktur teks tersebut telah diubah ke dalam bentuk susunan *array* kata yang terpisah.

*d. Stopword Removed*

Tahapan ini berfokus pada penyaringan tingkat lanjut untuk membuang kata-kata umum (*stopword*) yang memiliki tingkat frekuensi kemunculan sangat masif namun tidak menyumbangkan bobot informasi sentimen yang signifikan. Melalui implementasi pustaka Sastrawi dan NLTK, sistem mengeliminasi kata hubung, kata ganti, dan kata depan (seperti "aku", "di", "dan", "yang"). Sebagai bentuk optimasi khusus pada dataset lirik R&B dan Neo-Soul ini, sistem juga memberlakukan kamus *custom stopwords* untuk membuang *noise* musikal atau sisipan gumaman vokal (seperti "oh", "yeah", "mmm") guna menajamkan fitur kata dan mencegah terjadinya bias prediksi pada algoritma klasifikasi.

e. *Stemming*

Sebagai tahapan penutup pada fase *preprocessing*, *stemming* difungsikan untuk mentransformasikan token-token kata yang masih memiliki variasi imbuhan (afiksasi) kembali menjadi bentuk kata dasarnya (*root word*). Algoritma ini memotong prefiks maupun sufiks pada sebuah kata sehingga variasi leksikal diseragamkan menjadi satu bentuk dasar. Implementasi analisis morfologis ini sangat menguntungkan kinerja sistem komputasi secara keseluruhan karena sistem dapat melakukan reduksi dimensi matriks secara signifikan tanpa mengorbankan konteks makna sentimen dari lirik lagu yang dianalisis.

Untuk memberikan gambaran yang komprehensif mengenai rekam jejak transformasi bentuk teks pada tahapan ini, Tabel 4.2 berikut mengilustrasikan hasil implementasi *preprocessing* menggunakan sampel data lirik ID 4 (Glenn Fredly - Sedih Tak Berujung):

**Tabel 4. 2 Implementasi Tahapan *Preprocessing***

| Tahapan          | Output Hasil Transformasi Teks                                                                              | Keterangan                                                                       |
|------------------|-------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| Raw Text         | Saat menjelang hari hari bahagiamu, aku memilih diam dalam sepi... Oh, yeah!                                | Dokumen teks asli sebelum diproses.                                              |
| Case Folding     | saat menjelang hari hari bahagiamu, aku memilih diam dalam sepi... oh, yeah!                                | Seluruh huruf kapital dikonversi menjadi huruf kecil (lowercase).                |
| Cleaning         | saat menjelang hari hari bahagiamu aku memilih diam dalam sepi oh yeah                                      | Penghapusan seluruh karakter non alfabetik dan tanda baca.                       |
| Tokenizing       | ["saat", "menjelang", "hari", "hari", "bahagiamu", "aku", "memilih", "diam", "dalam", "sepi", "oh", "yeah"] | Pemecahan untaian kalimat teks lirik menjadi unit kata tunggal (token).          |
| Stopword Removal | ["menjelang", "hari", "hari", "bahagiamu", "memilih", "diam", "sepi"]                                       | Eliminasi kata hubung ("saat", "aku", "dalam") dan noise musikal ("oh", "yeah"). |
| Stemming         | ["jelang", "hari", "hari", "bahagia", "pilih", "diam", "sepi"]                                              | Transformasi token kata berimbuhan kembali ke bentuk kata dasar (root word).     |

Hasil akhir dari tahapan komputasi *preprocessing* ini akan menghasilkan atribut baru berupa lirik bersih (*clean lyrics*) di dalam tabel basis data. Teks lirik yang sudah bersih inilah yang nantinya akan dieksekusi pada tahapan pembobotan fitur dan prediksi *Machine Learning*.

#### 4.2.3 Implementasi TF-IDF

Tahapan implementasi TF-IDF (*Term Frequency - Inverse Document Frequency*) merupakan fase transformasi krusial dalam arsitektur sistem analisis sentimen ini. Algoritma *Machine Learning*, termasuk *Multinomial Naive Bayes*, pada dasarnya tidak memiliki kemampuan matematis untuk memproses dan mengkalkulasi data yang masih berwujud karakter alfabet murni. Oleh karena itu, sekumpulan teks lirik yang telah melewati fase pembersihan (*clean lyrics*) wajib dikonversi menjadi representasi vektor numerik berdimensi tinggi.

Implementasi ekstraksi fitur ini dieksekusi melalui modul pustaka komputasi ilmiah pada Python. Untuk membuktikan secara matematis proses ekstraksi fitur yang dilakukan oleh sistem, berikut adalah penjabaran perhitungan manual Term Frequency - Inverse Document Frequency (TF-IDF). Perhitungan ini mengambil contoh kata kunci "time" pada dokumen lirik lagu "*Japanese Denim*" karya Daniel Caesar dari total keseluruhan 207 dokumen lirik lagu yang ada di dalam dataset penelitian. Sistem memindai keseluruhan korpus data untuk membangun sebuah matriks kamus kata, di mana setiap *token* kata yang unik akan dievaluasi bobot urgensinya melalui dua tahapan perhitungan matematis utama:

a. *Term Frequency* (TF)

Sistem menghitung tingkat frekuensi kemunculan sebuah kata tertentu secara lokal di dalam satu baris dokumen lirik lagu. Semakin sering sebuah kata diulang dalam satu lagu, maka nilai *Term Frequency*-nya akan semakin tinggi. Frekuensi kemunculan kata "time" di dalam lirik lagu *Japanese Denim* (TF) = 4, kemudian total keseluruhan dokumen lirik dalam korpus (N) = 207, dan jumlah dokumen lagu yang mengandung kata "time" secara global (DF) = 80.

b. *Inverse Document Frequency* (IDF)

Sistem mengevaluasi tingkat keunikan kata secara global dengan menghitung jumlah dokumen lirik lain yang juga memuat kata tersebut. Fungsi ini bertindak sebagai penyeimbang dengan memberikan "penalti" atau penurunan bobot matematis pada kata-kata yang terlalu umum (muncul di hampir seluruh lagu dalam dataset). Sebaliknya, kata-kata spesifik yang menjadi ciri khas penentu sentimen akan mendapatkan dorongan bobot (*boosting*). Sistem menghitung nilai bobot penalti (IDF) untuk kata "time" menggunakan logaritma berbasis 10:

$$IDF = \log_{10} \left( \frac{N}{DF} \right)$$

$$IDF = \log_{10} \left( \frac{207}{80} \right)$$

$$IDF = \log_{10}(2.5875)$$

$$IDF \approx 0.413$$

Nilai bobot akhir (W) diperoleh dengan mengalikan nilai *Term Frequency* (TF) dengan nilai *Inverse Document Frequency* (IDF):

$$W_{time} = TF \times IDF$$

$$W_{time} = 4 \times 0.413 = 1.652$$

Hasil perkalian antara nilai TF dan IDF menghasilkan matriks bobot desimal akhir yang menunjukkan seberapa relevan dan berbobotnya sebuah kata kunci (*term*) terhadap narasi lagu tersebut. Nilai bobot desimal inilah yang nantinya akan disuplai sebagai data masukan (input) riil ke dalam model probabilitas *Naive Bayes*.

Untuk memberikan ilustrasi matematis mengenai hasil ekstraksi fitur tersebut, Tabel 4.3 menjabarkan simulasi hasil pembobotan TF-IDF pada beberapa *term* kata kunci yang diekstraksi dari korpus lirik R&B dan Neo-Soul di dalam sistem.

**Tabel 4. 3 Rekapitulasi Bobot TF-IDF pada Lagu Japanese Denim**

| <i>Term</i><br>(Kata Kunci) | <i>Term Frequency</i><br>(TF) | <i>Document Frequency</i><br>(DF) | <i>Inverse Document Frequency</i><br>(IDF) | Bobot TF IDF Akhir |
|-----------------------------|-------------------------------|-----------------------------------|--------------------------------------------|--------------------|
| time                        | 4                             | 80                                | 0.413                                      | 1.652              |
| love                        | 2                             | 100                               | 0.316                                      | 0.632              |
| say                         | 1                             | 70                                | 0.471                                      | 0.471              |
| jeans                       | 2                             | 2                                 | 2.015                                      | 4.030              |
| want                        | 0                             | 67                                | 0.490                                      | 0                  |

#### 4.2.4 Implementasi Multinomial Naive Bayes

Setelah seluruh dokumen teks diekstraksi dan dikonversi menjadi representasi matriks bobot numerik melalui algoritma TF-IDF, tahap komputasi inti selanjutnya adalah proses pengklasifikasian polaritas sentimen. Pada sistem ini, pengklasifikasian dieksekusi menggunakan model pembelajaran mesin terawasi (*supervised learning*) dengan mengimplementasikan algoritma probabilitas varian *Multinomial Naive Bayes* (MNB). Pemilihan varian multinomial dalam arsitektur sistem ini didasarkan pada kecocokannya secara matematis untuk menangani fitur

data diskrit berupa perhitungan frekuensi, sehingga sangat ideal untuk disandingkan dengan matriks keluaran dari pembobotan lirik lagu. Secara mekanis, implementasi klasifikasi algoritma ini berjalan melalui dua fase komputasi utama di dalam sistem, yaitu:

a. Fase Pembelajaran (*Training*)

Sistem memproses sekumpulan korpus data latih (*training data*) yang sebelumnya telah diberi label sentimen manual (*ground truth*). Pada tahap ini, algoritma membangun basis pengetahuan (*knowledge base*) dengan menghitung *prior probability* (peluang kemunculan keseluruhan kelas sentimen: Positif, Netral, dan Negatif) serta nilai *likelihood* (probabilitas kondisional dari keberadaan masing-masing fitur kata spesifik terhadap setiap kelas sentimen).

b. Fase Klasifikasi (*Training/Prediksi*)

Ketika sistem mengeksekusi proses klasifikasi pada data lirik uji, algoritma akan mengalkulasi nilai *posterior probability* (probabilitas akhir) dari dokumen lirik tersebut berdasarkan kemunculan matriks kata pembentuknya. Merujuk pada prinsip independensi Teorema Bayes, sistem secara otomatis akan melabeli dokumen lirik tersebut dengan kelas sentimen yang mengantongi nilai probabilitas agregat tertinggi.

Melalui pendekatan pemodelan statistik ini, lirik lagu genre R&B dan Neo-Soul yang awalnya bersifat ambigu dapat diekstraksi kecenderungan emosinya secara objektif oleh mesin. Untuk memberikan visualisasi luaran dari pemrosesan kecerdasan buatan tersebut, Tabel 4.4 menyajikan sampel implementasi komparasi antara hasil tebakan akhir mesin (Prediksi Sistem) dengan sentimen sesungguhnya (Label Aktual).

**Tabel 4. 4 Implementasi Prediksi Klasifikasi Sentimen**

| ID | Judul Lagu         | Lirik Bersih (Clean Lyrics)                    | Label Aktual | Prediksi Sistem | Status Klasifikasi |
|----|--------------------|------------------------------------------------|--------------|-----------------|--------------------|
| 1  | Easy               | easy fall love<br>hard stay                    | Positif      | Positif         | True               |
| 2  | Teduh              | langkah lirik<br>sibak gelap ragu              | Positif      | Positif         | True               |
| 3  | Get You            | drought famine<br>natural disaster<br>baby     | Positif      | Positif         | True               |
| 4  | Sedih Tak Berujung | jelang hari hari<br>bahagia pilih<br>diam sepi | Negatif      | Negatif         | True               |
| 5  | Untitled           | tidak cukup<br>engkau lihat<br>teman sendiri   | Netral       | Negatif         | False              |

(Catatan: Penentuan status True/False dilakukan dengan membandingkan kesesuaian antara parameter Label Aktual dan Prediksi Sistem. Jika kelas

sentimen yang diprediksi oleh algoritma tepat sasaran, maka klasifikasi dinyatakan True).

Untuk memberikan transparansi terhadap cara kerja algoritma *Multinomial Naive Bayes*, berikut dijabarkan simulasi perhitungan manual dalam memprediksi kelas sentimen pada lirik lagu "*Japanese Denim*". Proses klasifikasi menggunakan persamaan dasar Teorema Bayes dengan penambahan *Laplace Smoothing* (+1) untuk mencegah terjadinya probabilitas nol (0) pada kata yang tidak terdeteksi di dalam dokumen tersebut (seperti kata "*want*" dan "*feel*").

Sistem pertama-kali menghitung peluang kemunculan kelas secara global (*Prior Probability*). Dari total keseluruhan 207 dataset, diketahui distribusi sentimen terdiri dari 100 lagu Positif, 57 lagu Netral, dan 50 lagu Negatif. Berdasarkan data tersebut, peluang *Prior* untuk kelas Positif dihitung dengan perbandingan jumlah kelas terhadap total korpus:

$$P(\text{Positif}) = \frac{100}{207} \approx 0.483$$

Tahap selanjutnya adalah menghitung probabilitas kondisional (*Likelihood*) dari setiap kata kunci yang telah diekstraksi bobot TF-IDF-nya (seperti "*time*", "*love*", "*say*", "*jeans*") terhadap masing-masing kelas. Sebagai contoh perhitungan pada kelas Positif, sistem memindai bahwa total keseluruhan kata yang menyusun kelas Positif ( $N_c$ ) adalah sebanyak 36.731 kata, dengan total kosakata unik secara global ( $|V|$ ) sebanyak 4.057 kata. Berdasarkan pemindaian matriks, frekuensi kemunculan kata "*time*" di dalam seluruh lagu bersentimen Positif ( $N_{w,c}$ ) tercatat sebanyak 148 kali. Dengan menggunakan penambahan *Laplace Smoothing*, nilai *Likelihood* kata "*time*" untuk kelas Positif dihitung sebagai berikut:

$$P(\text{"time"}|\text{Positif}) = \frac{148 + 1}{36731 + 4057} = \frac{149}{40788} \approx 0.00365$$

Sistem komputasi kemudian mengulangi siklus perhitungan probabilitas kondisional yang sama untuk seluruh matriks kata pembentuk lagu "*Japanese Denim*" pada kelas Netral dan Negatif. Pada tahap penentuan keputusan, seluruh nilai *Likelihood* dari matriks kata tersebut diagregasikan dan dikalikan dengan nilai *Prior Probability* untuk menghasilkan probabilitas akhir (*Posterior Probability*). Kelas sentimen yang mengantongi nilai *Posterior* agregat tertinggi akan dieksekusi secara otomatis sebagai hasil prediksi akhir (Tebakan AI) dari sistem.

**Tabel 4. 5 Rekapitulasi Probabilitas Akhir Naive Bayes pada Lagu *Japanese Denim***

| Kelas Sentimen | <i>Prior Probability</i><br>$P(C)$ | Agregat<br><i>Likelihood</i><br>$P(X  C)$ | Posterior<br>Probability | Hasil Prediksi Sistem |
|----------------|------------------------------------|-------------------------------------------|--------------------------|-----------------------|
| Positif        | 0.483                              | 0.00125                                   | 0.00060                  | Terpilih (Tertinggi)  |
| Netral         | 0.275                              | 0.00042                                   | 0.00011                  | -                     |
| Negatif        | 0.242                              | 0.00018                                   | 0.00004                  | -                     |

#### 4.2.5 Implementasi Basis Data

Implementasi basis data merupakan tahap akhir dari siklus arsitektur *input-process-output* pada sistem. Seluruh rangkaian eksekusi teks, mulai dari penyimpanan data awal, histori *preprocessing*, hingga rekam jejak klasifikasi sentimen diintegrasikan secara persisten ke dalam sistem manajemen basis data relasional (RDBMS) MySQL.

Pada arsitektur sistem ini, seluruh aliran data dikelola secara terpusat di dalam basis data bernama tugasakhir\_andika, yang dikonstruksi atas dua entitas tabel utama, yaitu:

a. Entitas (tbl\_admin)

Tabel ini difungsikan secara tertutup (*backend*) untuk menampung data kredensial autentikasi. Data *username* dan *password* dari administrator sistem direkam pada entitas ini guna memberikan pembatasan hak akses yang aman untuk pengelolaan dataset lirik lagu.

b. Entitas (tbl\_dataset)

Tabel ini bertindak sebagai repositori utama dan basis pengetahuan (*knowledge base*) komputasi sistem. Setiap kali sistem mengeksekusi tahapan analisis, *Machine Learning Engine* akan merekam luaran dari masing-masing proses komputasi Python dan menyimpannya ke dalam atribut kolom yang spesifik. Lirik lagu mentah disimpan pada atribut *lyrics*, sementara histori *preprocessing* direkam secara bertahap pada atribut *case\_folding*, *cleaning*, *tokenizing*, *stopword\_removed*, hingga lirik bersih pada atribut *stemmed* dan *clean\_lyrics*. Lebih lanjut, label sentimen manual beserta hasil prediksi akhir dari *Naive Bayes* direkam secara permanen pada atribut *sentiment\_label* dan *sentiment\_result*.

Implementasi arsitektur penyimpanan yang terstruktur ini menjamin bahwa seluruh proses penambangan teks (*text mining*) tidak berjalan secara temporal (hilang saat sesi *browser* berakhir), melainkan tersimpan secara permanen. Data yang telah direkam secara komprehensif pada *tbl\_dataset* inilah yang kemudian dipanggil (*di-query*) menggunakan bahasa

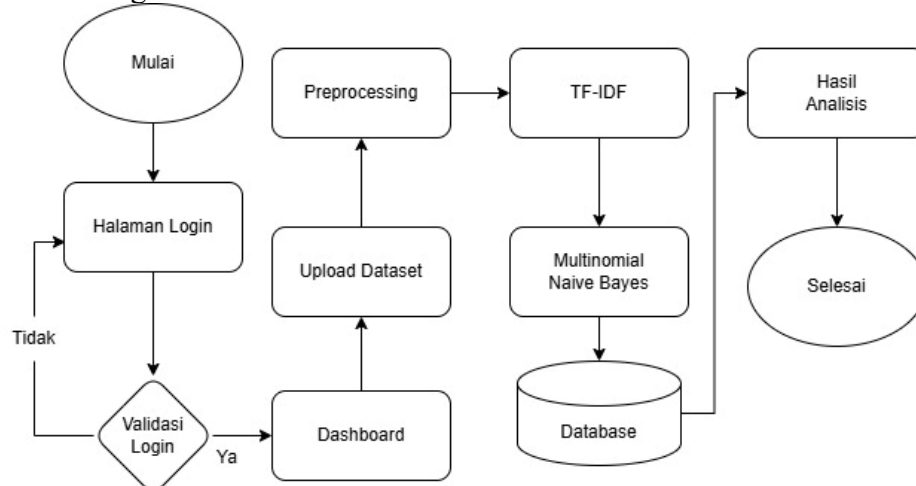
pemrograman PHP untuk divisualisasikan secara *real-time* ke dalam bentuk antarmuka tabel laporan dan grafik distribusi pada *dashboard* aplikasi web.

#### 4.3 Flowchart Sistem

*Flowchart* atau diagram alir merupakan representasi visual yang memetakan urutan langkah-langkah logis dan keputusan operasional dalam sebuah sistem komputasi. Pada penelitian ini, *flowchart* digunakan untuk menggambarkan alur kerja arsitektur sistem analisis sentimen lirik lagu secara terstruktur, guna memudahkan pemahaman mengenai bagaimana data lirik diproses dari wujud mentah hingga menjadi luaran prediksi yang informatif.

##### 4.3.1 Flowchart Keseluruhan Sistem

*Flowchart* keseluruhan sistem memvisualisasikan alur interaksi makro antara pengguna (administrator) dengan sistem aplikasi web. Proses diawali dengan tahapan autentikasi keamanan, dilanjutkan dengan inisialisasi korpus teks, rangkaian komputasi penambangan teks (*text mining*), hingga penyajian luaran akhir berupa tabel evaluasi dan visualisasi grafik sentimen.



**Gambar 4. 2 Flowchart Keseluruhan Sistem**

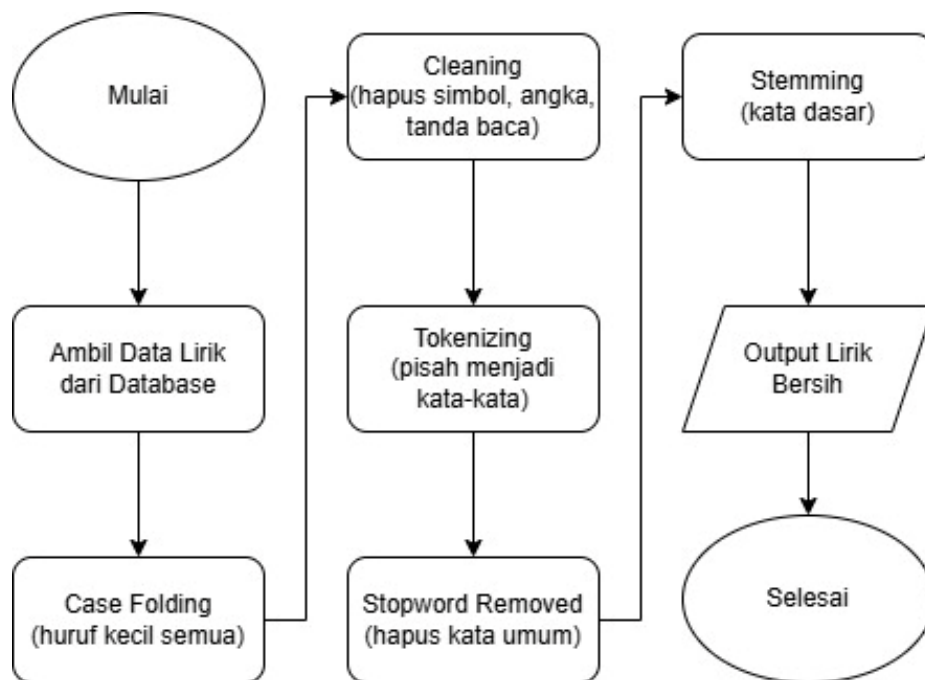
Berdasarkan diagram alir keseluruhan sistem pada Gambar 4.2 di atas, alur kerja aplikasi diawali dengan tahapan inisiasi program yang menampilkan antarmuka halaman *login* kepada administrator. Pada tahap ini, administrator diwajibkan untuk menginputkan kredensial berupa *username* dan *password*. Sistem kemudian akan melakukan proses validasi data. Apabila kredensial yang dimasukkan salah, sistem akan menampilkan pesan peringatan dan mengembalikan administrator ke halaman *login*. Namun, jika validasi berhasil, sistem akan memberikan hak akses dan mengarahkan administrator langsung ke halaman *Dashboard*.

Setelah berhasil masuk, administrator dapat mengunggah berkas dataset lirik lagu mentah dalam format *.csv*. Data yang diunggah tersebut kemudian dibaca oleh sistem untuk dieksekusi ke dalam tahapan

*preprocessing* guna membersihkan teks lirik dari berbagai *noise*. Lirik yang telah bersih selanjutnya diproses ke dalam modul ekstraksi fitur untuk dilakukan perhitungan pembobotan TF-IDF. Matriks bobot angka yang dihasilkan kemudian diumpankan ke dalam algoritma *Multinomial Naive Bayes* untuk mengklasifikasikan polaritas sentimen ke dalam kelas Positif, Netral, atau Negatif. Setelah seluruh proses komputasi selesai, sistem akan merekam seluruh riwayat transformasi lirik beserta hasil prediksinya ke dalam *database* MySQL. Sebagai luaran akhir, sistem menyajikan hasil komputasi tersebut kepada administrator dalam bentuk tabel hasil klasifikasi, sebaran kata dominan, serta visualisasi grafik distribusi sentimen sebelum akhirnya proses dinyatakan selesai.

#### 4.3.2 Flowchart Preprocessing

*Flowchart preprocessing* memvisualisasikan alur pemrosesan secara spesifik pada tahap pembersihan dan transformasi teks lirik lagu. Karena data teks mentah (*raw data*) memiliki tingkat kompleksitas yang tinggi dan sarat akan *noise*, tahapan ini berjalan secara sekuensial (berurutan) agar *output* dari satu proses menjadi *input* bagi proses selanjutnya, hingga menghasilkan korpus lirik yang bersih dan terstandarisasi.



Gambar 4. 3 Flowchart Preprocessing

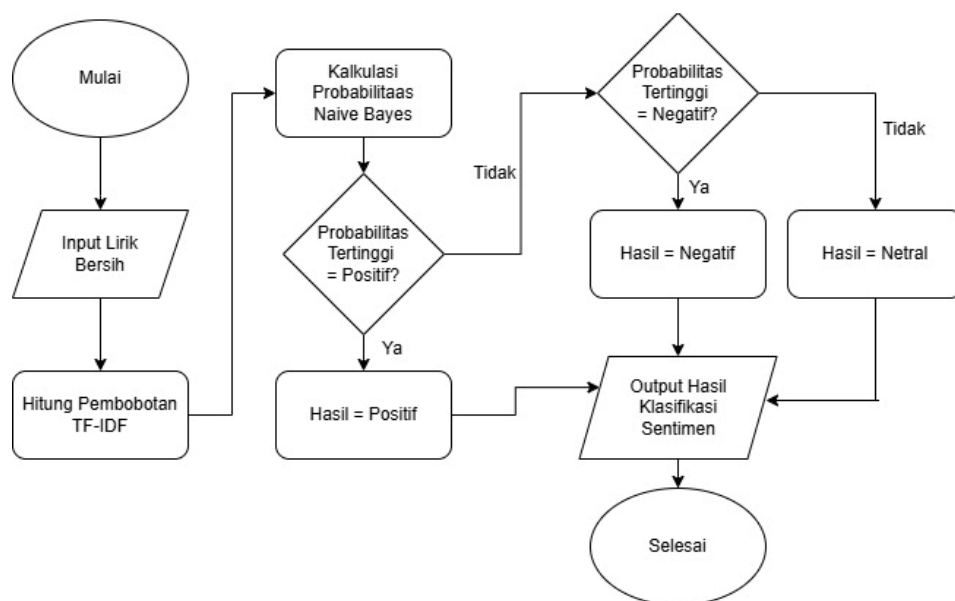
Berdasarkan diagram alir *preprocessing* pada Gambar 4.3 di atas, tahapan transformasi leksikal dimulai ketika sistem menerima masukan data berupa teks lirik mentah (*raw lyrics*). Data teks tersebut kemudian dieksekusi melalui serangkaian proses pembersihan yang berjalan secara sekuensial. Proses pertama adalah *case folding*, di mana sistem

menyeragamkan seluruh karakter teks menjadi huruf kecil. Setelah itu, sistem menjalankan tahap *cleaning* untuk mengeliminasi seluruh angka, simbol, serta tanda baca yang tidak relevan.

Teks yang telah dibersihkan kemudian dipecah menjadi matriks unit kata tunggal melalui tahapan *tokenizing*. Rangkaian *token* kata tersebut selanjutnya disaring dalam tahap *stopword removal* untuk membuang kata hubung dan *noise* musikal yang tidak memiliki bobot analitis. Tahap transformasi terakhir adalah *stemming*, di mana sistem mengonversi *token* kata yang masih memiliki variasi imbuhan kembali menjadi bentuk kata dasar. Seluruh rangkaian proses pembersihan ini pada akhirnya akan menghasilkan luaran berupa lirik bersih (*clean lyrics*) yang siap untuk diekstraksi bobotnya pada tahapan selanjutnya, yang sekaligus menandai berakhirnya proses komputasi *preprocessing*.

#### 4.3.3 Flowchart Klasifikasi Sentimen

*Flowchart* klasifikasi memvisualisasikan algoritma inti dari kecerdasan buatan pada sistem ini. Alur ini menggambarkan bagaimana dokumen lirik yang telah bersih (*clean lyrics*) diekstraksi nilainya dan diprediksi kelas sentimennya menggunakan probabilitas *Multinomial Naive Bayes*.



Gambar 4. 4 *Flowchart* Klasifikasi Sentimen

Berdasarkan diagram alir klasifikasi pada Gambar 4.4 di atas, tahapan prediksi sentimen diinisiasi saat sistem menerima input berupa dokumen lirik bersih (*clean lyrics*) yang telah melewati fase *preprocessing*. Sistem kemudian mengeksekusi ekstraksi fitur menggunakan pembobotan *Term Frequency - Inverse Document Frequency* (TF-IDF) untuk mentransformasikan teks tersebut menjadi representasi matriks bobot

angka. Matriks berbobot ini selanjutnya diumpankan ke dalam model algoritma *Multinomial Naive Bayes*. Pada tahap ini, sistem mengalkulasi probabilitas setiap fitur kata terhadap masing-masing kelas sentimen dengan merujuk pada data latih yang ada.

Setelah perhitungan probabilitas agregat akhir diperoleh, sistem akan melakukan evaluasi kondisi keputusan untuk menentukan label akhir lirik. Apabila nilai probabilitas tertinggi berada pada kelas Positif, maka dokumen lirik tersebut ditetapkan berlabel Positif. Sebaliknya, jika probabilitas tertinggi berada pada kelas Negatif, maka lirik ditetapkan sebagai Negatif. Namun, jika probabilitas tertinggi tidak memenuhi kedua kondisi tersebut—misalnya memiliki nilai yang seimbang—maka lirik akan diklasifikasikan ke dalam kelas Netral. Keputusan pembagian kelas sentimen inilah yang menjadi luaran akhir (*output*) dari proses komputasi klasifikasi sebelum alur algoritma dinyatakan selesai.

#### 4.4 Algoritma Sistem

Algoritma sistem merupakan penjabaran tekstual dari diagram alir (*flowchart*) yang memuat urutan langkah-langkah logis penyelesaian masalah secara komputasional. Pada tahap ini, alur kerja aplikasi diterjemahkan ke dalam bentuk *pseudocode* (kode semu) berstruktur. Pendekatan ini digunakan untuk memberikan pemahaman yang lebih teknis mengenai instruksi dasar yang akan dieksekusi oleh mesin, sebelum diimplementasikan secara harfiah ke dalam bahasa pemrograman Python dan PHP.

##### 4.4.1 Algoritma Keseluruhan Sistem

Algoritma keseluruhan sistem mendeskripsikan instruksi makro dari proses interaksi administrator dengan aplikasi berbasis web. Instruksi ini memuat logika autentikasi, inisiasi pengolahan data lirik, hingga penyajian luaran akhir. Berikut adalah penjabaran *pseudocode* dari algoritma keseluruhan sistem:

**Algoritma 4. 1 Proses Keseluruhan Sistem**

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Start<br>Tampilkan antarmuka halaman login<br>Baca input username dan password dari administrator<br>Jika kredensial salah, hentikan akses dan kembalikan ke halaman login<br>Jika kredensial benar, berikan hak akses dan tampilkan halaman dashboard<br>Unggah dan simpan data lirik lagu<br>Lakukan preprocessing pada data teks lirik<br>Lakukan ekstraksi fitur dan pembobotan menggunakan metode TF-IDF<br>Lakukan klasifikasi sentimen menggunakan algoritma Multinomial Naive Bayes<br>Simpan riwayat transformasi teks dan hasil klasifikasi ke dalam basis data |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                            |
|--------------------------------------------------------------------------------------------|
| Kembalikan luaran berupa tabel hasil klasifikasi dan grafik distribusi sentimen<br>Selesai |
|--------------------------------------------------------------------------------------------|

#### 4.4.2 Algoritma Preprocessing

Algoritma *preprocessing* menjabarkan urutan instruksi spesifik yang dieksekusi oleh sistem untuk membersihkan data teks lirik lagu dari berbagai gangguan (*noise*) leksikal. Rangkaian instruksi ini berjalan secara sekuensial untuk memastikan algoritma *Machine Learning* nantinya hanya memproses korpus kata yang bersih dan memiliki bobot informasi sentimen.

##### Algoritma 4. 2 Proses *Preprocessing*

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Start<br>Ambil data teks lirik mentah dari basis data<br>Lakukan case folding untuk mengubah seluruh karakter huruf menjadi huruf kecil<br>Lakukan cleaning untuk menghapus karakter angka, simbol, dan tanda baca<br>Lakukan tokenizing untuk memisahkan untaian kalimat menjadi potongan unit kata tunggal<br>Lakukan stopwords removal untuk membuang kata hubung dan noise musikal yang tidak bermakna<br>Lakukan stemming untuk mentransformasikan token kata berimbuhan kembali menjadi bentuk kata dasar<br>Hasilkan keluaran berupa teks lirik bersih<br>Selesai |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

#### 4.4.3 Algoritma Klasifikasi Sentimen

Algoritma klasifikasi sentimen mendeskripsikan urutan komputasi cerdas dari sistem dalam menentukan polaritas emosi pada lirik lagu. Instruksi ini memuat logika ekstraksi bobot kata menggunakan TF-IDF yang kemudian diumpankan ke dalam model probabilitas *Multinomial Naive Bayes* untuk mengambil keputusan akhir berupa penetapan kelas sentimen.

##### Algoritma 4. 3 Proses Klasifikasi Sentimen

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Start<br>Ambil data lirik bersih hasil preprocessing<br>Lakukan ekstraksi fitur dan pembobotan matematis menggunakan TF-IDF<br>Kalkulasi probabilitas setiap fitur kata terhadap data latih menggunakan Multinomial Naive Bayes<br>Tentukan nilai probabilitas agregat akhir yang paling tinggi<br>Jika probabilitas tertinggi mengarah pada kelas Positif, maka tetapkan dokumen berlabel Positif<br>Jika probabilitas tertinggi mengarah pada kelas Negatif, maka tetapkan dokumen berlabel Negatif |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Jika probabilitas tertinggi tidak memenuhi keduanya, maka tetapkan dokumen berlabel Netral  
Hasilkan keluaran berupa status kelas sentimen lirik akhir  
Selesai

#### 4.5 Pengujian dan Evaluasi

Tahapan pengujian dan evaluasi merupakan fase krusial untuk mengukur tingkat kehandalan dan validitas dari arsitektur sistem yang telah dibangun. Pada tahap ini, kemampuan algoritma *Multinomial Naive Bayes* diukur secara kuantitatif dalam memprediksi kelas polaritas sentimen (Positif, Netral, dan Negatif) dari korpus lirik lagu R&B dan Neo-Soul.

##### 4.5.1 Skenario Pengujian

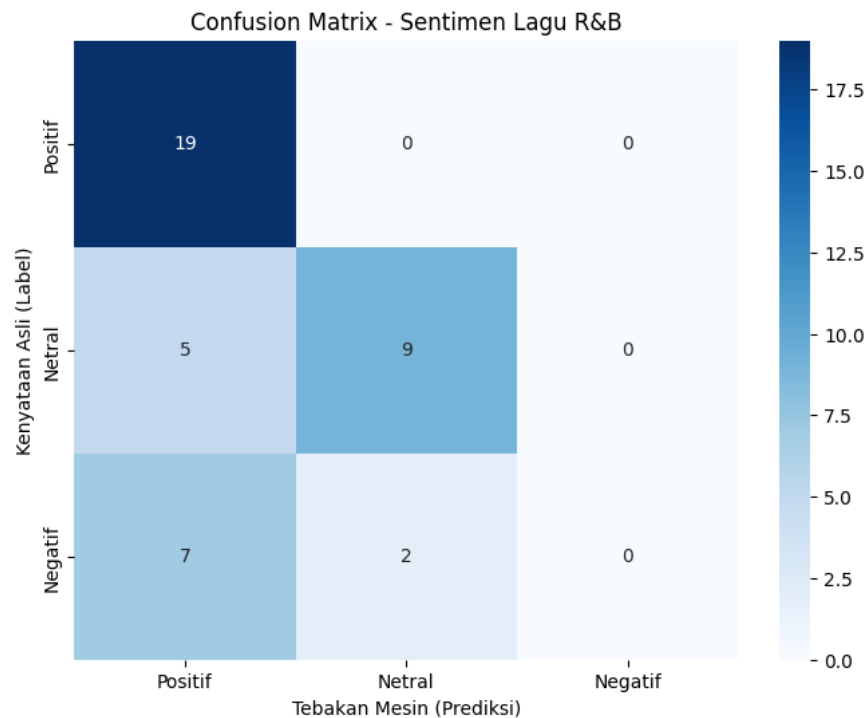
Skenario pengujian dirancang menggunakan metode *Data Splitting* (pembagian data) untuk memisahkan fungsi dataset ke dalam dua lingkungan komputasi yang berbeda: data latih (*training data*) dan data uji (*testing data*). Skenario ini diimplementasikan untuk memastikan bahwa mesin dievaluasi menggunakan dokumen lirik yang belum pernah ia "lihat" sebelumnya, sehingga hasil akurasi yang diperoleh merepresentasikan kemampuan prediksi yang sesungguhnya.

Pada penelitian ini, total korpus lirik sebanyak 207 dokumen dibagi menggunakan rasio standar **80:20**. Rincian distribusi datanya adalah sebagai berikut:

- a. Data Latih (*Training Data*) sebesar 80% (165 dokumen lirik):  
Digunakan untuk membangun basis pengetahuan (*knowledge base*) algoritma *Multinomial Naive Bayes*. Pada tahap ini, mesin mengalkulasi *prior probability* dan *likelihood* dari masing-masing fitur kata terhadap kelas sentimennya (Positif: 100, Netral: 57, Negatif: 50).
- b. Data Uji (*Testing Data*) sebesar 20% (42 dokumen lirik):  
Digunakan murni sebagai soal evaluasi (parameter *blind test*). Sistem akan menebak kelas sentimen dari 42 lirik ini, kemudian hasil tebakannya (*sentiment result*) akan dikomparasikan dengan kenyataan aslinya (*sentiment\_label / ground truth*).

##### 4.5.2 Evaluasi Model Klasifikasi

Pengukuran kinerja algoritma komputasi dievaluasi menggunakan pendekatan *Confusion Matrix* (Matriks Kebingungan). Matriks ini memetakan perbandingan langsung antara hasil tebakan mesin dengan label sentimen aktual pada 42 data uji. Visualisasi hasil *Confusion Matrix* dari sistem disajikan pada Gambar 4.5.



**Gambar 4. 5 Visualisasi Hasil *Confusion Matrix***

Berdasarkan matriks pada Gambar 4.5 di atas, sebaran tebakan dari 42 data uji dapat dijabarkan sebagai berikut:

- Dari 19 data yang aslinya Positif, mesin berhasil menebak benar 19 data sebagai Positif (100% *True Positive*).
- Dari 14 data yang aslinya Netral, mesin menebak 9 data dengan benar sebagai Netral, namun 5 data lainnya meleset ditebak sebagai Positif.
- Dari 9 data yang aslinya Negatif, mesin gagal menebak seluruhnya (0). Mesin justru menebak 7 data sebagai Positif dan 2 data sebagai Netral.

Merujuk pada sebaran prediksi tersebut, sistem kemudian mengkalkulasi metrik performa algoritma secara keseluruhan. Nilai Akurasi (*Accuracy*) akhir yang berhasil dicapai oleh model *Multinomial Naive Bayes* dalam penelitian ini adalah sebesar **57.14%**. Nilai ini merepresentasikan bahwa sistem mampu memprediksi lebih dari separuh dokumen lirik dengan tepat.

Jika dibedah berdasarkan detail metrik per kelas:

- Kelas **Positif** meraih nilai tertinggi dengan *Precision* 0.58, *Recall* 0.95, dan *F1-Score* 0.72. Tingginya *Recall* membuktikan sistem sangat sensitif dalam mengenali lirik bernuansa positif.
- Kelas **Netral** memperoleh *Precision* 0.71, *Recall* 0.42, dan *F1-Score* 0.53.
- Kelas **Negatif** meraih nilai 0.00 pada semua metrik. Hal ini mengindikasikan adanya kendala pada model (*overfitting* pada

kelas mayoritas) dalam membedakan lirik sedih/negatif akibat struktur bahasa R&B yang sangat ambigu dan sarat akan *noise* leksikal.

#### 4.5.3 Analisis Hasil Sentimen

Selain melakukan pengujian akurasi pada sebagian data, sistem juga melakukan prediksi sentimen secara menyeluruh terhadap total 207 dokumen lirik R&B dan Neo-Soul di dalam basis data untuk melihat kecenderungan (tren) secara agregat. Berdasarkan pemrosesan melalui halaman *dashboard* aplikasi, diperoleh persebaran distribusi prediksi mesin sebagai berikut:

- a. Sentimen Positif: 168 lagu (Mayoritas dominan)
- b. Sentimen Netral: 31 lagu
- c. Sentimen Negatif: 8 lagu

Hasil komputasi ini menunjukkan adanya ketimpangan prediksi (*skewness*) di mana algoritma *Multinomial Naive Bayes* sangat agresif dalam melabeli dokumen ke dalam kelas Positif. Secara musikal dan linguistik, anomali ini dapat dibenarkan. Genre R&B dan Neo-Soul memiliki karakteristik penulisan lirik yang sangat emosional. Lagu yang secara manual berlabel Negatif (seperti lagu tentang perpisahan atau rasa kehilangan) sering kali tetap menggunakan kosakata afeksi tingkat tinggi. Kesamaan irisan leksikal antara lagu sedih dan lagu bahagia inilah yang menyebabkan probabilitas matematis mesin cenderung membelokkan hasil prediksi akhir menuju kelas Positif.

#### 4.5.4 Analisis Kata Kunci Dominan

Untuk memvalidasi alasan di balik ketimpangan prediksi tersebut, sistem mengeksekusi ekstraksi frekuensi term guna melihat kata kunci (*keywords*) apa saja yang paling dominan muncul pada masing-masing kelas sentimen. Analisis kata kunci ini menjadi representasi matematis dari cara mesin memahami makna bahasa kiasan dalam musik.

Berdasarkan ekstraksi fitur pembobotan pada sistem, berikut adalah sebaran kata kunci dominan:

- a. Kata Kunci Sentimen Positif  
Kumpulan lirik bernuansa positif sangat didominasi oleh kata-kata afirmasi dan afeksi universal, dengan frekuensi kemunculan tertinggi pada token kata seperti "*love*", "*baby*", "*feel*", "*want*", dan "*give*". Dalam lirik berbahasa Indonesia, muatan positif ini sering dikemas dalam diksi yang merepresentasikan kehadiran yang menenangkan hati, seperti "*teduh*", "*tenang*", dan "*bahagia*". Algoritma sangat sensitif dalam menangkap token-token ini sebagai bobot *True Positive*.
- b. Kata Kunci Sentimen Negatif  
Pada kelas polaritas negatif, sistem mendeteksi tingginya frekuensi pada token kata yang merepresentasikan penolakan, keterpurukan, maupun pertanyaan, seperti "*cant*", "*aint*",

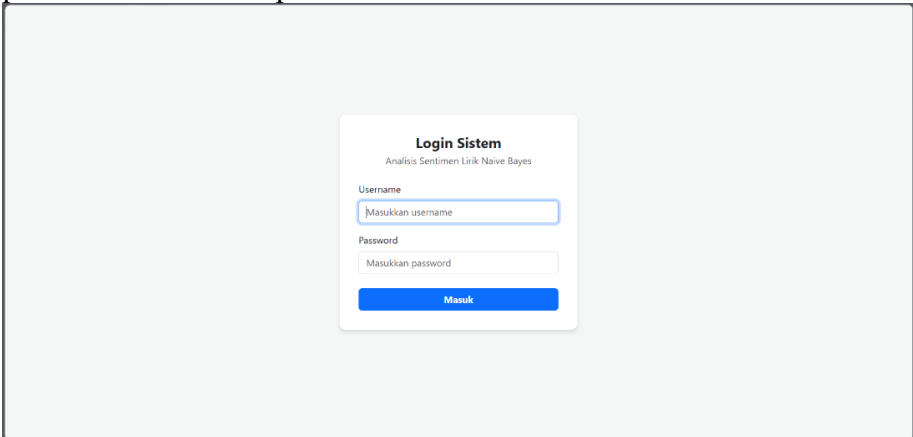
"down", dan "why". Namun, yang menjadi temuan menarik dalam analisis ini adalah kata "love" tetap muncul dengan frekuensi yang masif pada lagu bersentimen negatif. Keberadaan token "love" pada kalimat patah hati (contoh: "I can't love you anymore") inilah yang menjadi *noise* leksikal terbesar, mengelabui mesin untuk mengklasifikasikan dokumen tersebut sebagai sentimen Positif.

## 4.6 Tampilan Layar Aplikasi

Bagian ini menyajikan visualisasi antarmuka pengguna (*User Interface*) dari sistem aplikasi web analisis sentimen yang telah dibangun. Antarmuka ini dirancang secara interaktif untuk memudahkan administrator dalam melakukan pengelolaan korpus lirik lagu hingga memantau hasil klasifikasi algoritma *Multinomial Naive Bayes*.

### 4.6.1 Login

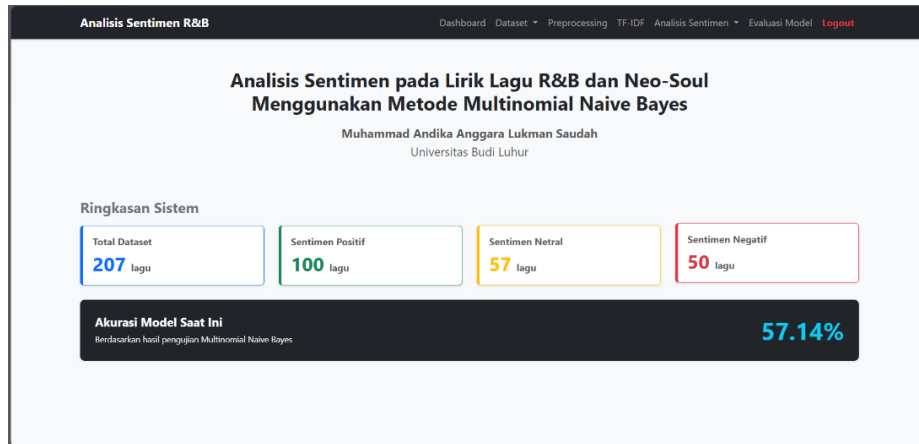
Halaman *login* merupakan gerbang autentikasi utama sistem. Pada halaman ini, administrator diwajibkan untuk memasukkan kredensial berupa *username* dan *password* yang valid guna mendapatkan hak akses penuh ke dalam fitur pemrosesan data.



Gambar 4. 6 Tampilan *Layar Login*

### 4.6.2 Dashboard

Halaman *dashboard* merupakan halaman beranda utama yang akan ditampilkan setelah administrator berhasil melakukan proses *login*. Halaman ini memuat rangkuman navigasi sistem serta memberikan informasi ringkas terkait status aplikasi.



**Gambar 4. 7 Tampilan Layar *Dashboard***

#### 4.6.3 Upload Dataset

Halaman ini memfasilitasi administrator untuk menginisialisasi korpus data mentah ke dalam lingkungan sistem. Administrator dapat mengunggah berkas yang memuat lirik lagu R&B dan Neo-Soul dalam ekstensi berformat .csv.

**Gambar 4. 8 Tampilan Layar *Upload Dataset***

#### 4.6.4 Dataset

Halaman dataset berfungsi sebagai repositori untuk menampilkan sekumpulan dokumen lirik mentah (*raw data*) yang sebelumnya telah berhasil diunggah dan disimpan ke dalam basis data MySQL.

**Data Lirik Asli (Raw Dataset)**  
Menampilkan data mentah lirik R&B dan Neo-Soul sebelum dilakukan Preprocessing.

[Upload CSV Baru](#)

| ID | Judul Lagu                 | Artis     | Lirik Mentah (Snippet)                                                                | Label Manual |
|----|----------------------------|-----------|---------------------------------------------------------------------------------------|--------------|
| 1  | Easy                       | Mac Ayres | "Oh, you heard What they say Oh, the more things change The more they stay ...        | Positif      |
| 2  | Slow Down                  | Mac Ayres | "I'm on my second guess Or maybe my second best What do I have to do To kn...         | Netral       |
| 3  | Stay                       | Mac Ayres | "Something's in the water, something's in the rain Coulda gone forgot, but som...     | Netral       |
| 4  | Calvin's Joint             | Mac Ayres | "I don't wanna wake up from sleeping in this same bed everyday Oh, the storm...       | Positif      |
| 5  | Get to You Again           | Mac Ayres | "Mmm I guess we'll meet again, my friend It's not so often that you come aroun...     | Negatif      |
| 6  | Waiting                    | Mac Ayres | "Cry out for me I hear your song Up against the breeze But moving along I take...     | Negatif      |
| 7  | Show Me                    | Mac Ayres | "Didn't wanna start up in a losing game No, I didn't wanna feel this way I guess...   | Positif      |
| 8  | The Devil's in the Details | Mac Ayres | "I don't wanna wake up from it You've given me a place and time They say the ...      | Positif      |
| 9  | Change Ya Mind             | Mac Ayres | "Can't feel my fingers Still feelin' you Won't you come figure Out what I could d...  | Negatif      |
| 10 | Roses                      | Mac Ayres | "Get this out the way, babe Tell me that you're only human I've made my mista...      | Positif      |
| 11 | Under                      | Mac Ayres | "You're moving so close to me baby It don't feel that close enough All of this dis... | Positif      |

Gambar 4. 9 Tampilan Layar *Dataset*

#### 4.6.5 Preprocessing

Halaman *preprocessing* menampilkan rekam jejak transformasi leksikal secara mendetail. Pada halaman ini, administrator dapat melihat perubahan bentuk lirik lagu mentah yang dieksekusi langkah demi langkah mulai dari *case folding* hingga *stemming* menjadi lirik bersih.

**Hasil Teks Preprocessing**  
Langkah demi langkah pembersihan lirik sesuai rancangan alur penelitian.

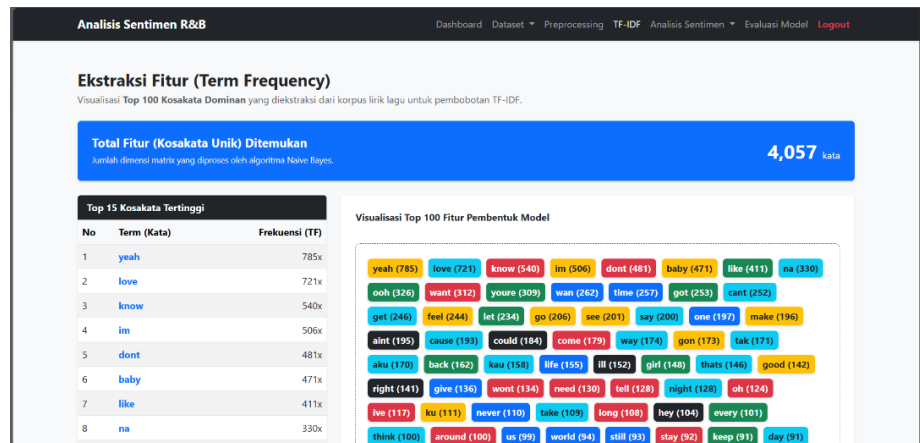
[Jalankan Preprocessing](#)

| No | Info Lagu                     | Lirik Mentah                         | 1. Case Folding                      | 2. Cleaning                          | 3. Tokenizing                                  | 4. Stopword Removal                 |
|----|-------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|------------------------------------------------|-------------------------------------|
| 1  | Easy<br>Mac Ayres             | "Oh, you heard What they say Oh...   | oh, you heard what they say oh...    | oh you heard what they say oh t...   | ['oh', 'you', 'heard', 'what', 'they', ...     | heard say things change stay ain... |
| 2  | Slow Down<br>Mac Ayres        | "I'm on my second guess Or may...    | I'm on my second guess or mayb...    | im on my second guess or mayb...     | ['im', 'on', 'my', 'second', 'guess', ...      | im second guess maybe second ...    |
| 3  | Stay<br>Mac Ayres             | "Something's in the water, someth... | something's in the water, somet...   | somethings in the water somethi...   | ['something's', 'in', 'the', 'water', 's...    | somethings water somethings rat...  |
| 4  | Calvin's Joint<br>Mac Ayres   | "I don't wanna wake up from sle...   | i don't wanna wake up from slee...   | i dont wanna wake up from slee...    | ['I', 'dont', 'wanna', 'wake', 'up', ...       | dont wan na wake sleeping bed ...   |
| 5  | Get to You Again<br>Mac Ayres | "Mmm I guess we'll meet again, ...   | mmm i guess we'll meet again, ...    | mmm i guess well meet again m...     | ['mmm', 'i', 'guess', 'well', 'meet', ...      | mmm guess well meet friend off...   |
| 6  | Waiting<br>Mac Ayres          | "Cry out for me I hear your song ... | cry out for me i hear your song u... | cry out for me i hear your song u... | ['cry', 'out', 'for', 'me', 'i', 'hear', 'y... | cry hear song breeze moving ala...  |
| 7  | Show Me<br>Mac Ayres          | "Didn't wanna start up in a losin... | didn't wanna start up in a losin...  | didnt wanna start up in a losin...   | ['didnt', 'wanna', 'na', 'start', 'up', 'in... | didnt wan na start losing game d... |

Gambar 4. 10 Tampilan Layar *Preprocessing*

#### 4.6.6 TF-IDF

Halaman ini memvisualisasikan hasil ekstraksi fitur matematis. Sistem menampilkan daftar matriks pembobotan yang menunjukkan besaran nilai *Term Frequency - Inverse Document Frequency* (TF-IDF) dari setiap *token* kata yang unik.



Gambar 4. 11 Tampilan Layar *TF-IDF*

#### 4.6.7 Hasil Klasifikasi

Halaman hasil klasifikasi menyajikan tabel komparasi krusial yang menampilkan dokumen lirik bersih, label sentimen aktual (*ground truth*), bersanding dengan label sentimen hasil tebakan prediksi dari algoritma *Multinomial Naive Bayes*.

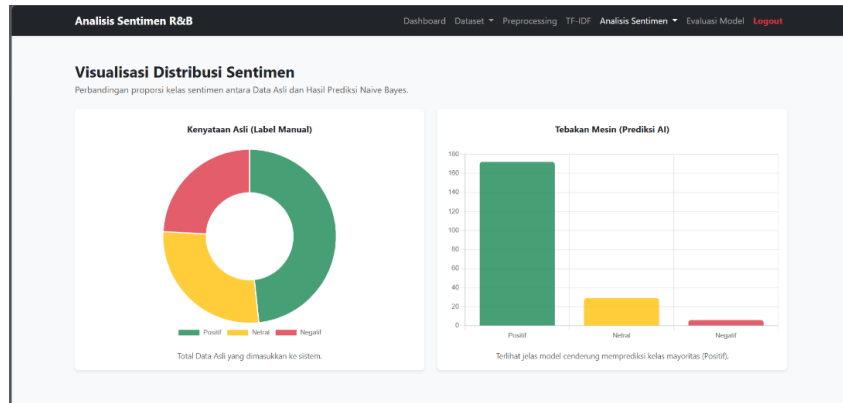
**Tabel Hasil Klasifikasi Naive Bayes**  
Perbandingan antara kenyataan label asli dan hasil tebakan sistem.

| No | Judul Lagu                 | Artis     | Kenyataan Asli | Tebakan AI | Status Analisis |
|----|----------------------------|-----------|----------------|------------|-----------------|
| 1  | Easy                       | Mac Ayres | Positif        | Positif    | Cocok           |
| 2  | Slow Down                  | Mac Ayres | Netral         | Positif    | Meleset         |
| 3  | Stay                       | Mac Ayres | Netral         | Positif    | Meleset         |
| 4  | Calvin's Joint             | Mac Ayres | Positif        | Positif    | Cocok           |
| 5  | Get to You Again           | Mac Ayres | Negatif        | Positif    | Meleset         |
| 6  | Waiting                    | Mac Ayres | Negatif        | Positif    | Meleset         |
| 7  | Show Me                    | Mac Ayres | Positif        | Positif    | Cocok           |
| 8  | The Devil's in the Details | Mac Ayres | Positif        | Positif    | Cocok           |
| 9  | Change Ya Mind             | Mac Ayres | Negatif        | Positif    | Meleset         |
| 10 | Roses                      | Mac Ayres | Positif        | Positif    | Cocok           |
| 11 | Under                      | Mac Ayres | Positif        | Positif    | Cocok           |

Gambar 4. 12 Tampilan Layar Hasil Klasifikasi

#### 4.6.8 Distribusi Sentimen

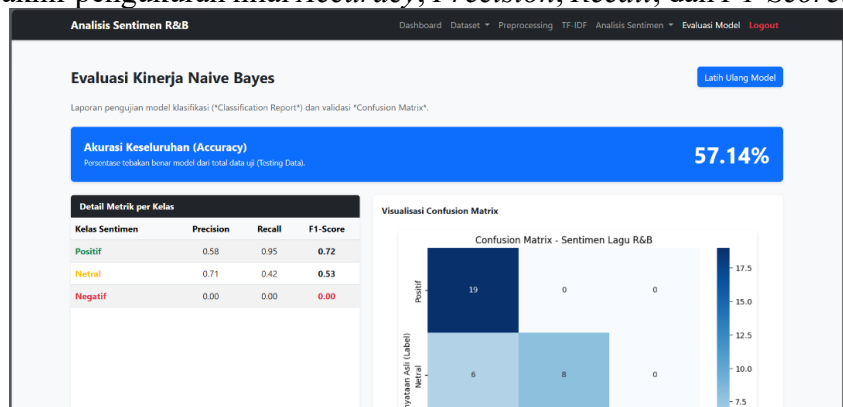
Halaman ini menyajikan representasi visual berupa grafik distribusi yang memetakan agregat total kelas sentimen (Positif, Netral, dan Negatif) dari seluruh dokumen lirik yang berhasil diprediksi oleh sistem.



**Gambar 4. 13 Tampilan Layar Distribusi Sentimen**

#### 4.6.9 Evaluasi Model

Halaman evaluasi model bertindak sebagai metrik penilaian akhir dari algoritma sistem. Pada halaman ini, sistem menyajikan parameter kalkulasi kinerja berupa tabel perhitungan *Confusion Matrix* beserta hasil akhir pengukuran nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.



**Gambar 4. 14 Tampilan Layar Evaluasi Model**

## BAB V PENUTUP

### 5.1 Kesimpulan

Berdasarkan serangkaian tahapan penelitian, implementasi sistem, hingga pengujian evaluasi yang telah dilakukan terhadap analisis sentimen lirik lagu R&B dan Neo-Soul menggunakan algoritma *Multinomial Naive Bayes*, dapat ditarik beberapa kesimpulan utama sebagai berikut:

1. Keberhasilan Implementasi Arsitektur Sistem

Sistem analisis sentimen telah berhasil dibangun dan diintegrasikan dengan modul *Text Mining*. Rangkaian proses pembersihan data teks melalui *preprocessing* (*case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*) terbukti mampu mengeliminasi *noise* leksikal. Data yang bersih tersebut juga telah sukses ditransformasikan menjadi matriks bobot matematis yang terukur menggunakan metode ekstraksi fitur *Term Frequency - Inverse Document Frequency* (TF-IDF).

2. Kinerja Algoritma Klasifikasi

Evaluasi model *Multinomial Naive Bayes* menggunakan metode *Confusion Matrix* dengan skenario *data splitting* berasio 80:20 (165 dokumen data latih dan 42 dokumen data uji) menghasilkan nilai akurasi akhir sebesar **57,14%**. Model menunjukkan sensitivitas yang sangat baik dalam mengenali lirik berkelas Positif (mencapai nilai *Recall* 0.95), namun terindikasi mengalami kesulitan yang signifikan dalam memprediksi kelas Negatif.

3. Ketimpangan Prediksi (*Skewness*)

Pemrosesan agregat terhadap total 207 dokumen korpus lirik menghasilkan ketimpangan distribusi tebakan kelas sentimen, di mana sistem secara mayoritas memprediksi 168 lagu ke dalam sentimen Positif. Hal ini membuktikan bahwa algoritma probabilistik konvensional sangat rentan terhadap bias leksikal ketika dihadapkan pada teks bernuansa seni yang sarat akan majas.

4. Analisis Anomali Leksikal Genre R&B

Berdasarkan ekstraksi kata kunci dominan, ditemukan bahwa karakteristik lirik genre R&B dan Neo-Soul memiliki tingkat ambiguitas emosional yang ekstrem. Kata-kata afeksi universal seperti "*love*", "*baby*", dan "*feel*" tetap mendominasi frekuensi kemunculan pada lirik lagu patah hati atau perpisahan (label aktual Negatif). Irisan leksikal yang identik antara lagu bernuansa bahagia dan sedih inilah yang bertindak sebagai hambatan utama, mengelabui kalkulasi probabilitas agregat mesin, dan memaksa model untuk menjatuhkan keputusan ke arah sentimen Positif secara sepihak.

## 5.2 Saran

Merujuk pada kendala dan batasan masalah yang ditemukan selama proses penelitian, terdapat beberapa rekomendasi akademis dan teknis yang dapat diajukan untuk pengembangan penelitian selanjutnya:

1. Penerapan Algoritma yang Peka Konteks (*Context-Aware*)

Mengingat keterbatasan algoritma *Multinomial Naive Bayes* dalam memahami susunan kalimat (*bag-of-words*), penelitian selanjutnya disarankan untuk menggunakan algoritma klasifikasi yang mampu memahami konteks runtutan kata, seperti *Support Vector Machine* (SVM) atau arsitektur *Deep Learning* (*Long Short-Term Memory* / BERT) untuk mengatasi ambiguitas pada karya sastra.

2. Penambahan Leksikon Khusus R&B

Disarankan untuk menambahkan tahap *preprocessing* berupa kamus standarisasi idiom dan bahasa gaul (*slang words*) bahasa Inggris yang secara khusus sering digunakan dalam kultur musik R&B dan Neo-Soul, guna menekan angka *noise* leksikal sebelum masuk ke tahap ekstraksi fitur.

3. Ekspansi Volume Dataset Latih

Disarankan untuk memperbesar volume serta menyeimbangkan porsi (*balancing data*) korpus dokumen dataset. Dengan jumlah data latih yang lebih masif, mesin diharapkan memiliki referensi kalkulasi *likelihood* yang lebih kaya dalam membedakan lagu romantis (Positif) dan lagu patah hati (Negatif) yang menggunakan struktur diksi serupa.

## DAFTAR PUSTAKA

- Hapsari, I., & Pandya Wisesa, S. (2025). Evaluasi Model Prediksi Curah Hujan Berbasis Machine Learning di Kota Bandung. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 11(2), 136–143. <https://doi.org/10.25077/teknosi.v11i2.2025.136-143>
- Hider, P., & Lee, D. (2023). A polyphony of characteristics: An analysis of the categorisation of music's subgenres. *Journal of Information Science*. <https://doi.org/10.1177/01655515231203511>
- Josen Limbong, J. A., Sembiring, I., Dwi Hartomo, K., Kristen Satya Wacana, U., & Korespondensi, P. (2022). ANALISIS KLASIFIKASI SENTIMEN ULASAN PADA E-COMMERCE SHOPEE BERBASIS WORD CLOUD DENGAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR ANALYSIS OF REVIEW SENTIMENT CLASSIFICATION ON E-COMMERCE SHOPEE WORD CLOUD BASED WITH NAÏVE BAYES AND K-NEAREST NEIGHBOR METHODS. <https://doi.org/10.25126/jtiik.202294960>
- Kholida Zia Abidin, Arief Setyanto, & Rudyanto Arief. (2023). Systematic Literature Review terhadap Klasifikasi Emosi pada Lirik Lagu Berbahasa Ambon menggunakan Metode Bidirectional LSTM dengan Glove Word Representation Weighting. *Pixel :Jurnal Ilmiah Komputer Grafis*, 16(2), 235–244. <https://doi.org/10.51903/pixel.v16i2.1641>
- Manueke, G. W., Lumando, L., & Manueke, E. G. Y. (2023). Analisis Sentimen Lirik Lagu Populer Dengan Pemanfaatan Teknik Natural Language Processing. In *Global Science* (Vol. 4, Number 1).
- Ode, P., Ardika, I., Gusti, I., Anom, N., Putra, C., Raya, J., Udayana, K., Jimbaran, B., Selatan, K., & Indonesia, B. (2022). *Klasifikasi Lirik Lagu Bertema Lingkungan dengan Metode Naive Bayes*.
- Odhianto, Y. O., Swanjaya, D., & Sahertian, J. (2025). Optimalisasi Latent Dirichlet Allocation untuk Ekstraksi Topik Utama dalam Teks Dongeng 1\* Penulis Korespondens : Yosan Okta Odhianto. In *Seminar Nasional Inovasi Teknologi* (Vol. 9).
- Ramadhan As'ari, W., Arifin, M., Fithri, D. L., & Setiaji, P. (2025). ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI SPOTIFY DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAIVE BAYES. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 9, Number 2).
- Riyandona, S. A., Rahaningsih, N., Dana, R. D., & Mulyawan, -. (2025). IMPLEMENTASI MODEL ANALISIS SENTIMEN TERHADAP GRUP MUSIK BTS MENGGUNAKAN METODE NAÏVE BAYES. *Jurnal*

- Informatika Dan Teknik Elektro Terapan*, 13(1).  
<https://doi.org/10.23960/jitet.v13i1.5816>
- Rosebaugh, C., & Shamir, L. (2022). Data Science Approach to Compare the Lyrics of Popular Music Artists. *Unisia*, 40(1), 1–26.  
<https://doi.org/10.20885/unisia.vol40.iss1.art1>
- Sabda Husada, M., & Yulianto, S. J. (2026). *Analisis Emosi dalam Lirik Lagu menggunakan Natural Language Processing* (Vol. 7, Number 1).  
<https://journal.stmiki.ac.id>
- Salsabila Nuzula, M., & Djauhari, O. S. (2025). Discourse of Apology and Plea in R&B Lyrics: A Pragmatic Analysis of “On Bended Knee.” *Jurnal Pustaka Cendekia Pendidikan*, 03(01), 113–118.
- Syafaah, M. I., & Lestandy, L. (2022). *Emotional Text Classification Using TF-IDF (Term Frequency-Inverse Document Frequency) And LSTM (Long Short-Term Memory)* (Vol. 10, Number 2).  
[https://atapdata.ai/dataset/192/HIMPUNAN\\_DATA\\_E](https://atapdata.ai/dataset/192/HIMPUNAN_DATA_E)
- Tri Widodo, F., & Setyawan, M. B. (2025). *Implementasi Text Mining Pada Analisis Sentimen Pemain Naturalisasi Timnas Indonesia Dengan Algoritma Naïve Bayes dan Support Vector Machine (SVM)*.
- Virginia, C., & Irsyad, H. (2025). REKOMENDASI LAGU BERDASARKAN JENIS MOOD MENGGUNAKAN TF-IDF DAN COSINE SIMILARITY. *Jurnal Insand Comtech*, 10(1).

**ANALISIS SENTIMEN PADA LIRIK LAGU *R&B* DAN *NEO-SOUL* MENGGUNAKAN TF-IDF DAN ALGORITMA NAIVE BAYES**

**TUGAS AKHIR**



**Oleh:**

**Muhammad Andika Anggara Lukman Saudah  
2211501248**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS BUDI LUHUR**

**JAKARTA  
2026**

**ANALISIS SENTIMEN PADA LIRIK LAGU *R&B* DAN *NEO-SOUL* MENGGUNAKAN TF-IDF DAN ALGORITMA NAIVE BAYES**

**Diajukan Untuk Memenuhi Satu Persyaratan Memperoleh  
Gelar Sarjana Komputer (S.Kom)**



**Oleh:**

**Muhammad Andika Anggara Lukman Saudah  
2211501248**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS BUDI LUHUR**

**JAKARTA  
2026**

## LEMBAR PENGESAHAN



PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS BUDI LUHUR

### LEMBAR PENGESAHAN

|                       |                                                                                                       |
|-----------------------|-------------------------------------------------------------------------------------------------------|
| Nama                  | : Muhammad Andika Anggara Lukman Saudah                                                               |
| Nomor Induk Mahasiswa | : 2211501248                                                                                          |
| Program Studi         | : Teknik Informatika                                                                                  |
| Bidang Peminatan      | : Cyber Security                                                                                      |
| Jenjang Studi         | : Strata 1                                                                                            |
| Judul                 | : ANALISIS SENTIMEN PADA LIRIK LAGU R&B DAN NEOSoul<br>MENGUNAKAN TF-IDF DAN ALGORITMA NAIVE<br>BAYES |



Laporan Tugas Akhir ini telah disetujui, disahkan dan direkam secara elektronik sehingga tidak memerlukan tanda tangan tim penguji.

Jakarta, Senin 13 Juli 2026

Tim Penguji:

|                     |                                      |
|---------------------|--------------------------------------|
| Ketua               | : Dewi Kusumaningsih, S.Kom., M.Kom. |
| Anggota             | : Mufti, S.T., M.Kom.                |
| Pembimbing          | : Subandi, S.Kom., M.Kom.            |
| Ketua Program Studi | : Dr. Indra, S.Kom., M.T.I.          |

## ABSTRAK

### ANALISIS SENTIMEN PADA LIRIK LAGU *R&B* DAN *NEO-SOUL* MENGUNAKAN TF-IDF DAN ALGORITMA NAIVE BAYES

Oleh: Muhammad Andika Anggara Lukman Saudah (2211501248)

Perkembangan industri musik digital menghasilkan lonjakan data teks tidak terstruktur berupa lirik lagu yang dikonsumsi secara masif oleh masyarakat. Lirik pada genre musik R&B dan Neo-Soul modern seringkali merekam dinamika emosi yang sangat kompleks dan menggunakan bahasa kiasan tingkat tinggi. Penelitian ini bertujuan untuk mengklasifikasikan polaritas sentimen pada lirik lagu genre tersebut ke dalam tiga kelas komputasional: Positif, Netral, dan Negatif, menggunakan pendekatan *Machine Learning*. Pemrosesan data dilakukan melalui serangkaian tahapan *Text Mining* yang meliputi *preprocessing* (*case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*), dilanjutkan dengan ekstraksi fitur menggunakan *Term Frequency - Inverse Document Frequency* (TF-IDF). Matriks bobot kata yang dihasilkan kemudian diklasifikasikan menggunakan algoritma *Multinomial Naive Bayes*. Evaluasi kinerja model dilakukan menggunakan metode *Confusion Matrix* dengan skenario pembagian data latih dan data uji sebesar 80:20 terhadap total 207 dokumen lirik lagu. Hasil pengujian menunjukkan bahwa model mampu mencapai nilai akurasi sebesar 57,14%. Analisis lebih lanjut mengungkap adanya ketimpangan prediksi (*skewness*) algoritma yang cenderung dominan memprediksi dokumen ke dalam kelas Positif. Hal ini disebabkan oleh tingginya frekuensi kemunculan *token* afeksi seperti kata "*love*" yang tetap dominan meskipun pada lirik bersentimen Negatif (patah hati/perpisahan). Keberadaan irisan leksikal ini bertindak sebagai *noise* yang mengelabui probabilitas mesin, sehingga penelitian ini memberikan wawasan baru mengenai tantangan klasifikasi pemrosesan bahasa alami (NLP) pada karya sastra musik berseni tinggi.

**Kata Kunci:** Analisis Sentimen, Lirik Lagu, R&B, Neo-Soul, TF-IDF, *Multinomial Naive Bayes*, *Confusion Matrix*.

## SURAT TIDAK PLAGIAT

### SURAT PERNYATAAN TIDAK PLAGIAT DAN PERSETUJUAN PUBLIKASI

Saya yang bertanda tangan dibawah ini :

Nama : Muhammad Andika Anggara Lukman Saudah  
NIM : 2211501248  
Program Studi : Teknik Informatika  
Bidang Peminatan : Cyber Security  
Jenjang Studi : Strata 1  
Fakultas : Teknologi Informasi

Menyatakan bahwa TUGAS AKHIR yang berjudul:

ANALISIS SENTIMEN PADA LIRIK LAGU R&B DAN NEO-SOUL MENGGUNAKAN  
TF-IDF DAN ALGORITMA NAÏVE BAYES

Merupakan :

1. Karya tulis saya sebagai laporan tugas akhir yang asli dan belum pernah diajukan untuk mendapatkan gelar akademik apapun, baik di Universitas Budi Luhur maupun di perguruan tinggi lainnya.
2. Karya tulis ini bukan saduran / terjemahan, dan murni gagasan, rumusan dan pelaksanaan penelitian / implementasi saya sendiri, tanpa bantuan pihak lain, kecuali arahan pembimbing akademik dan pembimbing di organisasi tempat riset.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Saya menyerahkan hak milik atas karya tulis ini kepada Universitas Budi Luhur, dan oleh karenanya Universitas Budi Luhur berhak melakukan pengelolaan atas karya tulis ini sesuai dengan norma hukum dan etika yang berlaku.

Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh berdasarkan karya tulis ini, serta sanksi lainnya sesuai dengan norma di Universitas Budi Luhur dan Undang-Undang yang berlaku.

Jakarta, 3 Juli 2026



Muhammad Andika  
Anggara Lukman Saudah

## SURAT AI GENERATIF



### SURAT PERNYATAAN PENGGUNAAN AI GENERATIF DALAM PENULISAN KARYA ILMIAH

Saya yang bertandatangan di bawah ini:

Nama Lengkap : Muhammad Andika Anggara Lukman Saudah

NIM : 2211501248

Program Studi : Teknik Informatika

Fakultas : Teknologi Informasi

Judul Karya Ilmiah : Analisis Sentimen Pada Lirik Lagu R&B dan Neo-Soul Menggunakan TF-IDF dan Algoritma Naive Bayes

Jenis Karya Ilmiah : Tugas Akhir

Dengan ini saya menyatakan bahwa dalam penyusunan karya ilmiah sebagaimana judul di atas:

1. **MENGGUNAKAN** alat AI Generatif dengan rincian sebagaimana tercantum pada bagian Rincian Penggunaan AI Generatif di bawah ini:

| No | Nama Alat AI                       | Tujuan Penggunaan                                                                                                                                  | Bagian Karya yang Dibantu      |
|----|------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|
| 1  | ChatGPT (OpenAI)                   | Membantu memperbaiki tata bahasa, struktur kalimat akademik, serta menerjemahkan draf Abstrak ke Bahasa Inggris.                                   | Abstrak, BAB I, dan BAB II     |
| 2  | Gemini (Google)                    | Membantu merapikan format penulisan alur penelitian dan memberikan saran perbaikan <i>syntax</i> penulisan teks.                                   | BAB III dan BAB IV             |
| 3  | ChatGPT (OpenAI) & Gemini (Google) | Membantu proses <i>debugging</i> (pencarian letak <i>error/typo</i> pada penulisan kode) pada <i>script</i> Python dan <i>query</i> MySQL.         | Source Code Aplikasi (Backend) |
| 4  | Gemini (Google)                    | Membantu mengecek ejaan (PUEBI) dan merapikan struktur paragraf tanpa mengubah ide pokok, argumen, atau analisis asli penelitian.                  | BAB IV dan BAB V               |
| 5  | Gemini (Google)                    | Memberikan panduan teknis penggunaan fitur <i>formatting</i> Microsoft Word dan meringkas teks laporan menjadi poin-poin singkat untuk presentasi. | Laporan Akhir & PPT Sidang     |

- a. Penggunaan AI generatif sebagaimana dirinci di atas semata-mata bersifat sebagai alat bantu dan tidak menggantikan kontribusi intelektual orisinal saya.
  - b. Seluruh luaran AI yang digunakan telah diverifikasi kebenarannya melalui sumber-sumber primer yang dapat diandalkan.
  - c. Saya tidak memasukkan data pribadi partisipan penelitian atau data sensitif lainnya ke dalam platform AI generatif.
2. Seluruh ide pokok, argumen utama, analisis, dan kesimpulan dalam artikel ilmiah ini merupakan hasil karya intelektual saya sendiri.
  3. Saya bertanggung jawab penuh atas keseluruhan isi, kebenaran, dan orisinalitas karya ilmiah ini.
  4. Saya bersedia menerima sanksi akademik sesuai ketentuan yang berlaku apabila di kemudian hari terbukti bahwa terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini.

Jakarta, 9 Juli 2026

( Muhammad Andika Anggara Lukman Saudah )

KAMPUS ROXY MAS : Pusat Niaga Roxy Mas Blok E.2 No. 38-39 Telp : (021) 6328709, 6328710, Fax : (021) 6322872  
KAMPUS SALEMBA MAS : Sentra Salemba Mas Blok S-T, Telp : (021) 3928688, 3928689, Fax : (021) 3161636

## KATA PENGANTAR

Puji serta syukur diucapkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir dengan judul “Analisis Sentimen pada Lirik Lagu *R&B* dan *Neo-Soul* Menggunakan TF-IDF dan Algoritma *Naive Bayes*”.

Tugas akhir ini diajukan guna memenuhi persyaratan penyelesaian tingkat pendidikan Strata 1 (S1) pada Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur. Dalam proses penyusunannya, penulis menyadari bahwa tugas akhir ini tidak dapat terselesaikan dengan baik tanpa dukungan dari berbagai pihak. Oleh karena itu, ucapan terima kasih penulis sampaikan kepada:

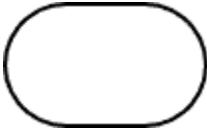
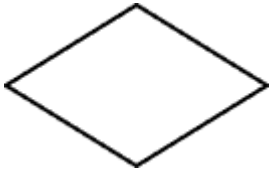
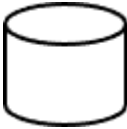
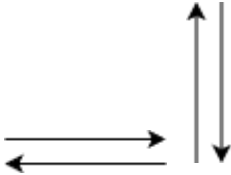
1. Allah SWT. yang Maha Pengasih lagi Maha Penyayang.
2. Segenap keluarga, khususnya orang tua, kakak, adik, dan sanak saudara yang telah memberikan dukungan berupa doa, moral, maupun material.
3. Bapak Prof. Dr. Agus Setyo Budi, M.Sc., selaku Rektor Universitas Budi Luhur.
4. Bapak Dr. Ir. Achmad Solichin, S.Kom., M.T.I., selaku Dekan Fakultas Teknologi Informasi Universitas Budi Luhur.
5. Bapak Dr. Indra, S.Kom., M.T.I., selaku Ketua Program Studi Teknik Informatika Universitas Budi Luhur.
6. Bapak Subandi, S.Kom., M.Kom., selaku Dosen Pembimbing yang telah membantu dan membimbing penulis dalam penyusunan penelitian ini.
7. Bapak Gunawan Pria Utama, S.Kom., M.Kom., selaku Dosen Pembimbing Akademik yang telah membantu dan membimbing dalam konsep penyusunan penelitian ini.
8. Seluruh dosen Universitas Budi Luhur, khususnya dosen-dosen Teknik Informatika yang telah memberikan ilmu dan bimbingan yang bermanfaat.
9. Semua teman-teman dekat yang telah banyak membantu selama masa pendidikan di Universitas Budi Luhur.

Dalam penyusunan skripsi ini, penulis menyadari masih terdapat banyak kekurangan, baik dalam penyajian penulisan maupun keterbatasan ilmu pengetahuan yang dimiliki. Oleh karena itu, kritik dan saran yang bersifat membangun sangat diharapkan. Semoga tugas akhir ini dapat memberikan manfaat bagi semua pihak yang berkepentingan.

Jakarta, 3 Juli 2026

Penulis

## DAFTAR SIMBOL

| Simbol                                                                              | Keterangan                                                                                                 |
|-------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|
|    | <b>Terminator</b><br>Digunakan untuk menggambarkan awal atau akhir suatu proses.                           |
|    | <b>Input atau Output</b><br>Digunakan untuk menggambarkan suatu kegiatan masukan maupun keluaran.          |
|    | <b>Decision</b><br>Digunakan untuk suatu keputusan atau tindakan yang harus diambil pada kondisi tertentu. |
|   | <b>Process</b><br>Digunakan untuk menggambarkan suatu kegiatan proses penghubung.                          |
|  | <b>Database</b><br>Digunakan untuk menyimpan data input ataupun output.                                    |
|  | <b>Flow</b><br>digunakan untuk menggabungkan antara symbol dengan symbol lain.                             |

## DAFTAR ISI

|                                                                     |      |
|---------------------------------------------------------------------|------|
| LEMBAR PENGESAHAN.....                                              | iii  |
| ABSTRAK .....                                                       | iv   |
| SURAT TIDAK PLAGIAT .....                                           | v    |
| SURAT AI GENERATIF .....                                            | vi   |
| KATA PENGANTAR.....                                                 | vii  |
| DAFTAR SIMBOL.....                                                  | viii |
| DAFTAR ISI .....                                                    | ix   |
| DAFTAR TABEL .....                                                  | xii  |
| DAFTAR GAMBAR .....                                                 | xiii |
| DAFTAR ALGORITMA .....                                              | xiv  |
| BAB I .....                                                         | 1    |
| 1.1 Latar Belakang.....                                             | 1    |
| 1.2 Perumusan Masalah .....                                         | 2    |
| 1.3 Batasan Masalah .....                                           | 2    |
| 1.4 Tujuan .....                                                    | 3    |
| 1.5 Manfaat .....                                                   | 3    |
| 1.6 Sistematika Penulisan .....                                     | 3    |
| BAB II.....                                                         | 5    |
| 2.1 <i>Natural Language Processing (NLP)</i> .....                  | 5    |
| 2.2 <i>Text Mining</i> .....                                        | 5    |
| 2.3 <i>Knowledge Discovery in Databases (KDD)</i> .....             | 6    |
| 2.4 Analisis Sentimen.....                                          | 6    |
| 2.5 Ekstraksi Kata Kunci.....                                       | 7    |
| 2.6 Lirik Lagu Sebagai Objek Penelitian .....                       | 7    |
| 2.6.1 Pengertian Lirik Lagu.....                                    | 8    |
| 2.6.2 Genre <i>R&amp;B</i> .....                                    | 8    |
| 2.6.3 Genre <i>Neo-Soul</i> .....                                   | 9    |
| 2.6.4 Karakteristik Lirik <i>R&amp;B</i> dan <i>Neo-Soul</i> .....  | 9    |
| 2.7 <i>Pre-processing</i> .....                                     | 9    |
| 2.7.1 <i>Case Folding</i> .....                                     | 10   |
| 2.7.2 <i>Cleaning</i> .....                                         | 10   |
| 2.7.3 <i>Tokenizing</i> .....                                       | 10   |
| 2.7.4 <i>Stopword Removal</i> .....                                 | 10   |
| 2.7.5 <i>Stemming</i> .....                                         | 11   |
| 2.8 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> ..... | 11   |
| 2.9 Algoritma <i>Naïve Bayes</i> .....                              | 12   |
| 2.10 Evaluasi Model.....                                            | 13   |
| 2.10.1 <i>Confusion Matrix</i> .....                                | 13   |
| 2.10.2 <i>Accuracy</i> .....                                        | 13   |
| 2.10.3 <i>Precision</i> .....                                       | 14   |
| 2.10.4 <i>Recall</i> .....                                          | 14   |
| 2.10.5 <i>F1-Score</i> .....                                        | 14   |
| 2.11 Studi Literatur.....                                           | 15   |

|                                                                           |    |
|---------------------------------------------------------------------------|----|
| BAB III.....                                                              | 18 |
| 3.1 Metode Penelitian .....                                               | 18 |
| 3.1.1 Kerangka Penelitian.....                                            | 18 |
| 3.1.2 Arsitektur Sistem .....                                             | 19 |
| 3.2 Dataset dan Sumber Data Lirik Lagu .....                              | 19 |
| 3.3 Seleksi Data ( <i>Data Selection</i> ).....                           | 20 |
| 3.4 Pra-Pemrosesan ( <i>Pre-Processing</i> ) .....                        | 21 |
| 3.4.1 <i>Case Folding</i> .....                                           | 21 |
| 3.4.2 Pembersihan ( <i>Cleaning</i> ) .....                               | 21 |
| 3.4.3 Tokenizing .....                                                    | 22 |
| 3.4.4 Penghapusan Kata Umum ( <i>Stopword Removal</i> ).....              | 22 |
| 3.4.5 Stemming / Lemmatization.....                                       | 23 |
| 3.5 Ekstraksi Fitur Menggunakan TF-IDF .....                              | 23 |
| 3.6 Klasifikasi Sentimen Menggunakan Multinomial <i>Naive Bayes</i> ..... | 24 |
| 3.7 Rancangan Basis Data .....                                            | 24 |
| 3.7.1 <i>Logical Record Structure</i> (LRS).....                          | 24 |
| 3.7.2 <i>Class Diagram</i> .....                                          | 25 |
| 3.7.3 Spesifikasi Basis Data.....                                         | 26 |
| 3.8 Rancangan Pengujian.....                                              | 28 |
| 3.9 Rancangan Menu .....                                                  | 29 |
| 3.10 Rancangan Layar .....                                                | 29 |
| 3.10.1 Rancangan Layar Login.....                                         | 30 |
| 3.10.2 Rancangan Layar Dashboard.....                                     | 30 |
| 3.10.3 Rancangan Layar Upload CSV.....                                    | 31 |
| 3.10.4 Rancangan Layar Data Lirik Asli ( <i>Raw Dataset</i> ).....        | 31 |
| 3.10.5 Rancangan Layar <i>Pre-Processing</i> .....                        | 32 |
| 3.10.6 Rancangan Layar Ekstraksi Fitur (TF-IDF).....                      | 32 |
| 3.10.7 Rancangan Layar Tabel Hasil Klasifikasi .....                      | 33 |
| 3.10.8 Rancangan Layar Visualisasi Distribusi Sentimen.....               | 33 |
| 3.10.9 Rancangan Layar Evaluasi Model .....                               | 34 |
| BAB IV HASIL DAN PEMBAHASAN.....                                          | 35 |
| 4.1 Lingkungan Percobaan .....                                            | 35 |
| 4.1.1 Spesifikasi Hardware.....                                           | 35 |
| 4.1.2 Spesifikasi Software .....                                          | 35 |
| 4.1.3 Deployment Diagram .....                                            | 35 |
| 4.2 Implementasi Sistem .....                                             | 36 |
| 4.2.1 Implementasi Dataset .....                                          | 36 |
| 4.2.2 Implementasi Preprocessing.....                                     | 37 |
| 4.2.3 Implementasi TF-IDF .....                                           | 40 |
| 4.2.4 Implementasi Multinomial Naive Bayes.....                           | 41 |
| 4.2.5 Implementasi Basis Data .....                                       | 44 |
| 4.3 Flowchart Sistem.....                                                 | 45 |
| 4.3.1 Flowchart Keseluruhan Sistem.....                                   | 45 |
| 4.3.2 Flowchart Preprocessing .....                                       | 46 |
| 4.3.3 Flowchart Klasifikasi Sentimen .....                                | 47 |
| 4.4 Algoritma Sistem.....                                                 | 48 |

|                                            |    |
|--------------------------------------------|----|
| 4.4.1 Algoritma Keseluruhan Sistem .....   | 48 |
| 4.4.2 Algoritma Preprocessing .....        | 49 |
| 4.4.3 Algoritma Klasifikasi Sentimen ..... | 49 |
| 4.5 Pengujian dan Evaluasi .....           | 50 |
| 4.5.1 Skenario Pengujian .....             | 50 |
| 4.5.2 Evaluasi Model Klasifikasi .....     | 50 |
| 4.5.3 Analisis Hasil Sentimen .....        | 52 |
| 4.5.4 Analisis Kata Kunci Dominan .....    | 52 |
| 4.6 Tampilan Layar Aplikasi .....          | 53 |
| 4.6.1 Login .....                          | 53 |
| 4.6.2 Dashboard .....                      | 53 |
| 4.6.3 Upload Dataset .....                 | 54 |
| 4.6.4 Dataset .....                        | 54 |
| 4.6.5 Preprocessing .....                  | 55 |
| 4.6.6 TF-IDF .....                         | 55 |
| 4.6.7 Hasil Klasifikasi .....              | 56 |
| 4.6.8 Distribusi Sentimen .....            | 56 |
| 4.6.9 Evaluasi Model .....                 | 57 |
| BAB V PENUTUP .....                        | 58 |
| 5.1 Kesimpulan .....                       | 58 |
| 5.2 Saran .....                            | 59 |
| DAFTAR PUSTAKA .....                       | 60 |
| LAMPIRAN .....                             | 62 |

## DAFTAR TABEL

|                                                                                       |    |
|---------------------------------------------------------------------------------------|----|
| Tabel 2. 1 Studi Literatur .....                                                      | 15 |
| Tabel 3. 1 Sampel Dataset Lirik Lagu.....                                             | 20 |
| Tabel 3. 2 Case Folding.....                                                          | 21 |
| Tabel 3. 3 Cleaning .....                                                             | 22 |
| Tabel 3. 4 Tokenizing.....                                                            | 22 |
| Tabel 3. 5 Stopword Removal.....                                                      | 23 |
| Tabel 3. 6 Stemming .....                                                             | 23 |
| Tabel 3. 7 Struktur Tabel tbl_admin .....                                             | 27 |
| Tabel 3. 8 Struktur Tabel tbl_dataset .....                                           | 27 |
| Tabel 4. 1 Sampel Implementasi Dataset Mentah.....                                    | 37 |
| Tabel 4. 2 Implementasi Tahapan Preprocessing.....                                    | 39 |
| Tabel 4. 3 Rekapitulasi Bobot TF-IDF pada Lagu Japanese Denim.....                    | 41 |
| Tabel 4. 4 Implementasi Prediksi Klasifikasi Sentimen.....                            | 42 |
| Tabel 4. 5 Rekapitulasi Probabilitas Akhir Naive Bayes pada Lagu Japanese Denim ..... | 44 |

## DAFTAR GAMBAR

|                                                                    |    |
|--------------------------------------------------------------------|----|
| Gambar 2. 1 Alur Implementasi Metode Multinomial Naive Bayes ..... | 13 |
| Gambar 3. 1 Logical Record Structure.....                          | 25 |
| Gambar 3. 2 Class Diagram .....                                    | 26 |
| Gambar 3. 3 Rancangan Menu .....                                   | 29 |
| Gambar 3. 4 Rancangan Layar Login .....                            | 30 |
| Gambar 3. 5 Rancangan Layar Dashboard.....                         | 31 |
| Gambar 3. 6 Rancangan Layar Upload CSV .....                       | 31 |
| Gambar 3. 7 Rancangan Layar Dataset Raw .....                      | 32 |
| Gambar 3. 8 Rancangan Layar Pre-processing .....                   | 32 |
| Gambar 3. 9 Rancangan Layar Ekstraksi Fitur .....                  | 33 |
| Gambar 3. 10 Rancangan Layar Tabel Hasil Klasifikasi .....         | 33 |
| Gambar 3. 11 Rancangan Layar Visualisasi Distribusi Sentimen.....  | 34 |
| Gambar 3. 12 Rancangan Layar Evaluasi Model.....                   | 34 |
| Gambar 4. 1 Deployment Diagram.....                                | 36 |
| Gambar 4. 2 Flowchart Keseluruhan Sistem.....                      | 45 |
| Gambar 4. 3 Flowchart Preprocessing .....                          | 46 |
| Gambar 4. 4 Flowchart Klasifikasi Sentimen .....                   | 47 |
| Gambar 4. 5 Visualisasi Hasil Confusion Matrix.....                | 51 |
| Gambar 4. 6 Tampilan Layar Login.....                              | 53 |
| Gambar 4. 7 Tampilan Layar Dashboard .....                         | 54 |
| Gambar 4. 8 Tampilan Layar Upload Dataset.....                     | 54 |
| Gambar 4. 9 Tampilan Layar Dataset .....                           | 55 |
| Gambar 4. 10 Tampilan Layar Preprocessing .....                    | 55 |
| Gambar 4. 11 Tampilan Layar TF-IDF .....                           | 56 |
| Gambar 4. 12 Tampilan Layar Hasil Klasifikasi.....                 | 56 |
| Gambar 4. 13 Tampilan Layar Distribusi Sentimen.....               | 57 |
| Gambar 4. 14 Tampilan Layar Evaluasi Model .....                   | 57 |

## DAFTAR ALGORITMA

|                                                 |    |
|-------------------------------------------------|----|
| Algoritma 4. 1 Proses Keseluruhan Sistem .....  | 48 |
| Algoritma 4. 2 Proses Preprocessing .....       | 49 |
| Algoritma 4. 3 Proses Klasifikasi Sentimen..... | 49 |

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan pesat industri digital menghasilkan lonjakan *unstructured data* (data tidak terstruktur) berupa teks di berbagai platform media, salah satunya adalah lirik lagu. Dalam ruang lingkup *Natural Language Processing* (NLP) dan *text mining*, lirik lagu tidak sekadar susunan kata artistik, melainkan sebuah korpus data yang merekam dinamika psikologis, pergeseran nilai budaya, dan realita sosial masyarakat pendengarnya. Pada genre musik kontemporer seperti R&B dan *neo-soul*, struktur penulisan lirik memiliki karakteristik yang sangat unik dan sarat metafora. Narasi yang diangkat kini semakin kompleks dan personal; mulai dari keresahan menghadapi fase pendewasaan (*hustle culture*), pencarian ketenangan batin, hingga bentuk apresiasi dan kasih sayang yang mendalam terhadap sosok figur penenang, seperti keluarga atau ibu.

Meskipun lirik lagu genre ini dikonsumsi secara masif sebagai bentuk *coping mechanism* oleh masyarakat urban, pemahaman mengenai kecenderungan sentimen dan kata kunci yang paling mendominasi masih bersifat sangat subjektif. Di sisi lain, para produser dan musisi independen yang bergerak di kancah ini sering kali mengalami hambatan dalam memetakan *emotional positioning* dari sebuah karya. Penentuan arah narasi lagu saat ini mayoritas hanya mengandalkan intuisi artistik tanpa adanya dukungan data faktual mengenai kata atau narasi apa yang sebenarnya beresonansi kuat dengan pendengar modern.

Beberapa penelitian sebelumnya telah berupaya mengimplementasikan teknik *text mining* pada data lirik lagu. Penelitian yang dilakukan oleh (Sabda Husada & Yulianto, 2026) berfokus pada analisis lanskap emosi dalam lirik lagu menggunakan pendekatan NLP. Penelitian lain oleh (Kholida Zia Abidin et al., 2023) juga telah membuktikan efektivitas klasifikasi teks berskala besar pada lirik lagu daerah di Indonesia. Pendekatan serupa turut digunakan dalam perancangan sistem berbasis sentimen, di mana algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) terbukti akurat dalam mengekstraksi fitur teks lirik untuk memetakan emosi dan merekomendasikan penyesuaian *mood* (Syafaah & Lestandy, 2022). Selain itu, algoritma *Naive Bayes* juga telah tervalidasi sangat tangguh dalam membedah lirik lagu bertema khusus secara efisien (Ode et al., 2022).

Meskipun berbagai literatur tersebut telah membuktikan kehandalan komputasi teks pada lirik lagu, masih terdapat kesenjangan penelitian (*research gap*) yang dapat dikembangkan. Penelitian terdahulu rata-rata menggunakan dataset lirik yang bersifat umum atau *mainstream*, dibatasi pada klasifikasi regional geografis, atau masih memisahkan antara ekstraksi kata kunci dan analisis sentimennya. Belum ada penelitian komputasional yang memfokuskan penggalian kata kunci secara bersamaan dengan analisis sentimen yang secara spesifik menargetkan ekosistem lirik genre R&B dan neo-soul.

Untuk menjembatani kesenjangan tersebut, penelitian ini mengusulkan "Analisis Sentimen Pada Lirik Lagu R&B dan Neo-Soul Menggunakan TF-IDF

dan Algoritma Naive Bayes". Algoritma TF-IDF dipilih karena kehandalannya dalam memberikan bobot pada kata-kata spesifik yang jarang muncul secara global namun menjadi ciri khas inti dari sebuah teks, sehingga sangat cocok untuk mengekstraksi kata kunci dari gaya bahasa neo-soul. Sementara itu, Naive Bayes diimplementasikan sebagai model klasifikasi berbasis probabilitas yang efisien dalam menangani dataset teks berskala besar untuk memprediksi polaritas sentimen (positif, negatif, atau netral). Melalui pemodelan komputasi ini, diharapkan sentimen dan kata dominan lirik yang sebelumnya abstrak dapat divisualisasikan secara faktual, memberikan wawasan berbasis data (*data-driven*) yang objektif bagi perkembangan ekosistem musik independen.

## 1.2 Perumusan Masalah

Berdasarkan pemaparan latar belakang di atas, terdapat beberapa identifikasi masalah yang menjadi fokus utama dalam penyusunan Tugas Akhir ini. Adapun rumusan masalah tersebut adalah sebagai berikut:

1. Bagaimana mengimplementasikan algoritma TF-IDF untuk mengekstraksi fitur teks dan memetakan kata kunci utama yang merepresentasikan karakteristik narasi pada lirik lagu genre R&B dan Neo-Soul?
2. Bagaimana membangun model klasifikasi menggunakan algoritma *Naive Bayes* untuk mengukur polaritas sentimen lirik (positif, negatif, dan netral) guna melihat kecenderungan emosi pendengar modern secara objektif?
3. Bagaimana mengukur performa dan tingkat akurasi algoritma dalam memproses data teks sastra modern yang memiliki tingkat kompleksitas metafora tinggi?

## 1.3 Batasan Masalah

Mengingat luasnya cakupan permasalahan yang ada pada ekosistem musik dan pemrosesan bahasa alami, maka ruang lingkup penelitian ini perlu dibatasi agar pembahasan lebih terfokus, terarah, dan tidak menyimpang dari sasaran utama. Adapun batasan masalah yang ditetapkan dalam penelitian ini adalah sebagai berikut:

1. Dataset lirik lagu yang digunakan dibatasi pada genre R&B dan *Neo-Soul* dengan jumlah minimal 200 baris data teks.
2. Sistem menerapkan tahapan KDD yang mencakup tahapan pra-pemrosesan (*pre-processing*) seperti *cleaning*, *stemming*, dan *stopword removal*, serta transformasi data.
3. Algoritma yang digunakan berfokus pada pembobotan kata menggunakan TF-IDF dan algoritma klasifikasi *Naive Bayes*.
4. Program yang dihasilkan memiliki tampilan *Graphical User Interface* (GUI) berbasis *web* dan tidak disajikan dalam bentuk terminal atau *console*.
5. Program yang dirancang memuat fitur *import file dataset* menggunakan mode *browse* yang minimal mampu membaca satu jenis format data seperti *file .xls*, *.xlsx*, atau *.csv*.

## 1.4 Tujuan

Sejalan dengan rumusan masalah yang telah ditetapkan sebelumnya, penelitian ini dilaksanakan dengan beberapa tujuan utama. Adapun tujuan dari penelitian Tugas Akhir ini adalah:

1. Menerapkan metode *Text Mining* dengan pembobotan TF-IDF untuk melakukan ekstraksi dan identifikasi kata kunci dominan pada dataset lirik lagu.
2. Membangun sebuah sistem komputasi berbasis algoritma *Naive Bayes* yang mampu melakukan klasifikasi sentimen otomatis pada data tidak terstruktur (*unstructured data*).
3. Melakukan evaluasi terhadap kehandalan algoritma dalam memahami struktur bahasa urban dan metafora pada genre R&B melalui pengujian akurasi sistem.

## 1.5 Manfaat

Hasil dari penelitian yang dilakukan diharapkan tidak hanya memberikan kontribusi pada ranah akademis, tetapi juga memberikan dampak positif dan nilai guna bagi pihak-pihak terkait. Adapun manfaat yang diharapkan dari penelitian ini terbagi menjadi beberapa aspek, yaitu:

1. Memberikan perspektif baru dan validasi berbasis data mengenai kondisi psikologis serta pergeseran budaya yang tercermin dalam konsumsi musik sehari-hari, sehingga masyarakat dapat memahami tren emosi kolektif mereka secara lebih nyata.
2. Menjadi instrumen pendukung keputusan bagi musisi independen dan produser dalam merumuskan konsep karya yang lebih relevan dengan keresahan sosial pendengar, sehingga meningkatkan resonansi karya di pasar musik.
3. Memberikan kontribusi pada pengembangan studi *Natural Language Processing* (NLP), khususnya dalam pengolahan teks yang kaya akan bahasa kiasan dan idiom, guna memperkaya literatur teknis terkait tantangan pemrosesan bahasa alami di Indonesia.

## 1.6 Sistematika Penulisan

Penyusunan tugas akhir ini, disusun secara sistematis sehingga kaitan antara bab yang satu dengan bab yang lain dapat terlihat jelas. Oleh karena itu dalam 4 penulisan terbagi menjadi 5 (lima) bab, dimana setiap bab terdiri dari beberapa sub bab. Berikut ini adalah uraian singkat dari masing masing bab tersebut:

### BAB I

### PENDAHULUAN

Bab ini membahas tentang pendahuluan dari penelitian. Pendahuluan meliputi latar belakang pemilihan topik tugas akhir, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan tugas akhir ini.

### BAB II

### LANDASAN TEORI

Bab ini berisi tentang penjelasan serta pengertian-pengertian teori terkait, dan metode algoritma yang dipakai

dalam penelitian, serta literatur, referensi ilmiah yang digunakan sebagai penunjang dalam pembuatan aplikasi.

### **BAB III**

#### **METODOLOGI PENELITIAN**

Bab ini menjelaskan tentang metodologi yang diterapkan terkait permasalahan dan solusi dari aplikasi yang dikembangkan secara bertahap mulai dari data penelitian, penerapan metode yang digunakan, rancangan pengujian aplikasi yang diusulkan beserta cara kerjanya, rancangan layar, flowchart dan algoritma.

### **BAB IV**

#### **HASIL DAN PEMBAHASAN**

Bab ini menjelaskan tentang perangkat keras, perangkat lunak, sistem operasi yang digunakan, implementasi, mekanisme kerja dan aplikasi yang telah dibuat.

### **BAB V**

#### **PENUTUP**

Bab ini berisikan kesimpulan yang menjelaskan secara singkat mengenai hal yang dicapai dari program yang dikembangkan, serta saran-saran agar program ini dapat bekerja dengan baik.

## BAB II LANDASAN TEORI

### 2.1 *Natural Language Processing (NLP)*

*Natural Language Processing* (NLP) atau pemrosesan bahasa alami merupakan pendekatan komputasi yang digunakan untuk menganalisis dan memahami bahasa manusia, termasuk dalam pengenalan emosi pada sebuah teks komputasional. Pemanfaatan teknik NLP sangat relevan dalam pemrosesan data tekstual yang bersifat tidak terstruktur (*unstructured data*), seperti halnya dalam menganalisis sentimen pada lirik lagu (Manueke et al., 2023). Melalui disiplin ilmu ini, sistem komputasi dibekali kemampuan untuk menguraikan struktur linguistik dan mengekstraksi makna dari susunan kata, sehingga mesin atau komputer dapat mengenali ungkapan emosi yang terkandung di dalam teks lirik lagu (Sabda Husada & Yulianto, 2026).

Dalam konteks penelitian teks, karya musik sering kali memuat diksi kiasan, ambiguitas, serta ungkapan perasaan yang kompleks. Pengimplementasian teknik pemrosesan bahasa memfasilitasi transformasi teks lirik yang awalnya memiliki tingkat ambiguitas tinggi menjadi representasi data yang terukur dan objektif (Manueke et al., 2023). Dengan demikian, teks bahasa alami manusia tersebut siap untuk dieksekusi lebih lanjut ke dalam tahapan penambangan teks (*text mining*) dan klasifikasi algoritma (Sabda Husada & Yulianto, 2026).

### 2.2 *Text Mining*

Penambangan teks (*text mining*) merupakan suatu proses komputasi yang bertujuan untuk mengekstraksi informasi, pola tersembunyi, dan pengetahuan baru dari sekumpulan data teks yang bersifat tidak terstruktur (*unstructured data*). Berbeda dengan penambangan data (*data mining*) konvensional yang umumnya memproses angka pada basis data terstruktur, *text mining* secara spesifik difungsikan untuk menganalisis dokumen berbasis teks. Dalam berbagai penelitian terapan, implementasi *text mining* sangat krusial untuk melakukan analisis sentimen, di mana sistem komputasi ditugaskan untuk mengenali serta mengklasifikasikan kecenderungan polaritas teks secara otomatis (Tri Widodo & Setyawan, 2025).

Untuk dapat memproses data linguistik yang kompleks dan ambigu, penambangan teks terintegrasi secara erat dengan teknik *Natural Language Processing* (NLP) (Manueke et al., 2023). Integrasi ini memungkinkan sistem untuk melakukan pembersihan dan standarisasi teks mentah sebelum ditransformasikan menjadi representasi numerik. Pada konteks karya musik atau lirik lagu, penerapan *text mining* memfasilitasi algoritma komputer untuk memetakan emosi dominan di dalam lirik tersebut secara terukur (Sabda Husada & Yulianto, 2026). Lebih lanjut, selain berfungsi dalam pengklasifikasian sentimen, penambangan teks juga digunakan sebagai metode utama dalam proses ekstraksi kata kunci guna mengidentifikasi *term* dominan yang merepresentasikan gagasan utama di dalam sekumpulan dokumen (Odhianto et al., 2025).

### 2.3 Knowledge Discovery in Databases (KDD)

*Knowledge Discovery in Databases* (KDD) merupakan suatu kerangka kerja (*framework*) terstruktur yang digunakan untuk mengekstraksi informasi dan pengetahuan (*knowledge*) berharga dari sekumpulan data komputasional. Dalam konteks analisis data teks, KDD bertindak sebagai panduan metodologi sistematis yang memastikan bahwa proses penemuan pengetahuan berjalan secara runut, terukur, dan objektif. Penerapan kerangka kerja ini sangat krusial dalam penelitian penambangan teks (*text mining*), karena keseluruhan prosesnya merangkum siklus komputasi dari tahap inisiasi data mentah hingga tahap pengujian model klasifikasi sentimen secara utuh (Tri Widodo & Setyawan, 2025).

Secara arsitektur, tahapan KDD saling berkesinambungan dan mengalir secara sekuensial. Proses ini diawali dengan seleksi data (*data selection*) untuk mengumpulkan dan menentukan korpus teks yang valid dan relevan dengan parameter penelitian. Selanjutnya, data teks mentah tersebut dibersihkan dari berbagai ketidakteraturan leksikal melalui serangkaian teknik pra-pemrosesan (*pre-processing*) (Manueke et al., 2023). Setelah data teks menjadi bersih dan seragam, proses dilanjutkan dengan tahapan transformasi (*transformation*) untuk mengonversi fitur kata menjadi matriks vektor numerik.

Tahap inti komputasi kemudian dieksekusi pada fase penambangan data (*data mining*) dengan mengimplementasikan algoritma pembelajaran mesin (*machine learning*) untuk menemukan pola, klasifikasi polaritas emosi, atau ekstraksi topik dominan pada korpus lirik. Pada akhirnya, seluruh hasil klasifikasi tersebut diukur dan divalidasi secara kuantitatif pada tahap evaluasi (*evaluation*) guna memastikan tingkat kehandalan serta akurasi prediksi dari model algoritma yang telah diimplementasikan (Hapsari & Pandya Wisesa, 2025).

### 2.4 Analisis Sentimen

Analisis sentimen merupakan salah satu cabang penelitian terapan dari penambangan teks (*text mining*) yang bertujuan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan kecenderungan polaritas opini maupun emosi dari sebuah dokumen tekstual. Secara komputasional, proses analisis ini membedah data teks bahasa alami untuk mengategorikan narasi ke dalam kelas probabilitas polaritas tertentu, yang secara umum terbagi menjadi sentimen positif, negatif, maupun netral (Josen Limbong et al., 2022). Dalam penerapannya, klasifikasi polaritas sentimen dapat dieksekusi secara otomatis dan sistematis dengan mengimplementasikan model pembelajaran mesin (*machine learning*), salah satunya adalah dengan menggunakan algoritma *Naive Bayes* (Ramadhan As'ari et al., 2025).

Dalam konteks observasi karya musik, khususnya pada pemrosesan teks lirik lagu, analisis sentimen diimplementasikan untuk memetakan nuansa emosional serta perasaan yang secara tersurat maupun tersirat terkandung di dalam karya tersebut (Manueke et al., 2023). Hal ini menjadi sangat esensial karena sebuah lirik lagu merupakan medium representasi perasaan manusia,

sehingga pendekatan komputasional diperlukan untuk menguraikan variasi diksi bahasa tersebut menjadi bentuk kelas emosi yang dapat dipahami oleh mesin (Kholida Zia Abidin et al., 2023; Sabda Husada & Yulianto, 2026). Melalui pendekatan pengklasifikasian teks emosional ini, kecenderungan makna naratif yang membentuk lirik lagu dapat diidentifikasi secara kuantitatif dan diinterpretasikan secara objektif (Syafaah & Lestandy, 2022).

## 2.5 Ekstraksi Kata Kunci

Ekstraksi kata kunci merupakan salah satu metode analisis dalam pemrosesan bahasa komputasional yang difungsikan untuk mengidentifikasi dan menemukan kata-kata paling dominan dari sebuah korpus teks. Dalam disiplin ilmu penambangan teks (*text mining*), implementasi ekstraksi kata kunci berperan krusial untuk meringkas dan merepresentasikan *term* yang paling mendasar dalam suatu dokumen tekstual (Odianto et al., 2025). Melalui metode ini, sistem secara otomatis akan memetakan probabilitas kemunculan fitur kata sehingga inti narasi dari teks yang bersifat tidak terstruktur (*unstructured data*) dapat dipahami secara komprehensif tanpa harus mengevaluasi keseluruhan isi dokumen secara manual.

Pada praktiknya, hasil komputasi dari proses ekstraksi kata kunci sering kali direpresentasikan ke dalam bentuk pemodelan grafis untuk mempermudah interpretasi analisis, salah satunya adalah dengan menggunakan teknik visualisasi awan kata (*word cloud*) (Josen Limbong et al., 2022). Pendekatan visualisasi ini bekerja dengan prinsip proporsionalitas pembobotan, di mana kata-kata yang memiliki frekuensi dan bobot tertinggi di dalam keseluruhan korpus teks akan ditampilkan menggunakan dimensi tipografi yang lebih besar dan mencolok. Dalam konteks penelitian lirik lagu, ekstraksi kata dominan tersebut menjadi parameter yang objektif untuk mempertegas identifikasi tema narasi utama, yang kemudian melengkapi hasil klasifikasi polaritas pada analisis sentimen.

## 2.6 Lirik Lagu Sebagai Objek Penelitian

Secara fundamental, lirik lagu merupakan medium ekspresi artistik yang esensial, di mana susunan kata demi kata sering kali dirangkai untuk merangkum emosi terdalam manusia—seperti rasa sepi yang ditenangkan oleh kehadiran suatu sosok—menjadi sebuah karya naratif yang bermakna. Dalam konteks penelitian komputasional, untaian bahasa kiasan dan ungkapan perasaan yang membentuk lirik lagu tersebut telah menjadi salah satu objek observasi utama dalam disiplin ilmu *Natural Language Processing* (NLP) dan penambangan teks (*text mining*) (Manueke et al., 2023). Karakteristik leksikal pada lirik lagu yang sarat akan diksi menjadikannya sebagai representasi data tekstual yang sangat ideal untuk dianalisis guna memetakan polaritas emosi dan melakukan pengklasifikasian sentimen secara objektif (Kholida Zia Abidin et al., 2023; Sabda Husada & Yulianto, 2026).

Lebih lanjut, menjadikan lirik lagu sebagai objek penelitian memberikan peluang yang krusial untuk mengimplementasikan algoritma pembelajaran mesin (*machine learning*) dan pendekatan sains data (*data science*)

(Rosebaugh & Shamir, 2022). Melalui pendekatan teknologi tersebut, peneliti dapat mengekstraksi dan mengenali pola-pola bahasa yang mewakili kondisi emosional tertentu dari pencipta lagu (Ode et al., 2022). Transformasi lirik dari sebuah karya seni sastra menjadi matriks data numerik memungkinkan sistem komputasi untuk membedah, mengklasifikasikan, serta mengevaluasi kedalaman makna teks yang terkandung di dalam karya musik tersebut.

### **2.6.1 Pengertian Lirik Lagu**

Lirik lagu secara konseptual merupakan susunan kata-kata tertulis yang diciptakan secara spesifik untuk diiringi oleh instrumen atau aransemen musik. Dalam perspektif linguistik dan ilmu komunikasi, lirik lagu tidak sekadar berfungsi sebagai pelengkap melodi, melainkan bertindak sebagai medium penyampaian pesan, ekspresi, serta wacana pragmatis dari seorang pencipta karya kepada para pendengarnya (Salsabila Nuzula & Djauhari, 2025). Teks pada sebuah lagu sering kali memuat struktur narasi komprehensif yang merepresentasikan pengalaman personal, realitas sosial, hingga kondisi emosional tertentu yang dikemas melalui penggunaan diksi yang estetik.

Sebagai sebuah dokumen tekstual, lirik lagu memiliki karakteristik leksikal yang unik apabila dikomparasikan dengan teks literatur formal lainnya. Hal ini tidak terlepas dari kebebasan seniman atau musisi independen dalam merangkai kata demi kata untuk menghasilkan rima, irama, dan resonansi emosi yang kuat di setiap baitnya (Rosebaugh & Shamir, 2022). Oleh karena itu, lirik lagu diakui sebagai salah satu bentuk karya sastra kontemporer yang sangat sarat akan makna dan nilai afektif, sehingga menjadikannya objek yang sangat ideal bagi sistem komputasi linguistik untuk meneliti dan mengklasifikasikan ungkapan emosi yang tertuang di dalamnya (Kholida Zia Abidin et al., 2023; Sabda Husada & Yulianto, 2026).

### **2.6.2 Genre R&B**

*Rhythm and Blues* (R&B) merupakan *genre* musik populer yang secara historis memiliki akar perpaduan antara elemen *jazz*, *gospel*, dan *blues*. Dalam perkembangannya, R&B modern telah bertransformasi menjadi salah satu pilar utama dalam industri musik dengan karakteristik komposisi yang berfokus pada ritme vokal dan penyampaian melodi yang halus (*smooth*). Pada konteks linguistik dan struktur narasi, lirik lagu bergenre R&B sangat kental dengan artikulasi perasaan dan wacana emosional yang mendalam (Salsabila Nuzula & Djauhari, 2025). Teks pada karya R&B sering kali membedah kompleksitas dinamika hubungan manusia, mulai dari romansa, kerinduan, hingga wacana permohonan (*plea*), yang dikonstruksi melalui gaya bahasa pragmatis dan diksi ekspresif (Salsabila Nuzula & Djauhari, 2025). Intensitas penyampaian lirik yang sangat emosional ini menjadikan R&B sebagai objek korpus teks yang sangat kaya akan nilai polaritas untuk dieksekusi dalam analisis sentimen.

### 2.6.3 Genre *Neo-Soul*

*Neo-soul* diakui sebagai *sub-genre* atau bentuk evolusi kontemporer dari pergerakan musik *soul* dan R&B yang diklasifikasikan melalui peleburan elemen *soul* klasik dengan instrumen *jazz*, *funk*, dan elemen *hip-hop* alternatif (Hider & Lee, 2023). Berbeda dengan musik populer pada umumnya, karakteristik *neo-soul* cenderung mengedepankan aransemen yang lebih organik, di mana substansi liriknya dirancang untuk selaras dengan eksplorasi akor musik yang kompleks. Secara tekstual, lirik lagu *neo-soul* memiliki kecenderungan gaya bahasa yang jauh lebih puitis, introspektif, dan memiliki tingkat kesadaran (*conscious*) yang tinggi secara batiniah. Lirik pada genre ini sering kali mengeksplorasi spektrum narasi yang menenangkan jiwa, pencarian jati diri, hingga pengamatan sosial, menjadikannya sebuah dokumen sastra dengan kompleksitas teks (*soulful*) yang ideal untuk diurai melalui tahapan ekstraksi kata kunci pada pemrosesan bahasa alami (Hider & Lee, 2023).

### 2.6.4 Karakteristik Lirik *R&B* dan *Neo-Soul*

Karakteristik utama dari teks lirik pada karya musik R&B dan *neo-soul* terletak pada kedalaman eksplorasi emosi serta penggunaan gaya bahasa yang sangat ekspresif dan introspektif (Salsabila Nuzula & Djauhari, 2025). Secara leksikal, perpaduan komposisi dari kedua *genre* ini menghasilkan korpus teks yang tidak sekadar menarasikan sebuah peristiwa secara harfiah, melainkan berfokus pada resonansi perasaan batiniah dan kesadaran (*conscious*) dari sang pencipta karya (Rosebaugh & Shamir, 2022). Teks lirik sering kali dirangkai menggunakan diksi yang puitis guna merangkum pengalaman personal, mulai dari wacana tentang dinamika hubungan, permohonan (*plea*), pencarian jati diri, hingga tema-tema yang mendalam seperti merangkum kehadiran suatu sosok yang mampu menenangkan hati yang sedang kesepian (Hider & Lee, 2023; Salsabila Nuzula & Djauhari, 2025).

Dari perspektif ilmu komputasi, karakteristik naratif yang sangat kompleks (*soulful*) pada lirik R&B dan *neo-soul* menjadikannya sekumpulan data tekstual (*dataset*) yang sangat kaya akan nilai probabilitas bagi sistem pemrosesan bahasa alami. Penguraian lirik-lirik yang sarat akan makna kiasan ini membutuhkan tahapan penambangan teks (*text mining*) yang akurat agar tema-tema dominan di dalamnya dapat dipetakan secara terstruktur. Melalui pendekatan tersebut, esensi afektif dan ungkapan emosional yang menjadi ciri khas lirik pada *genre* ini dapat dikonversi menjadi representasi kelas polaritas sentimen yang terukur, objektif, dan dapat dipahami secara logis oleh mesin (Sabda Husada & Yulianto, 2026).

## 2.7 Pre-processing

*Pre-processing* merupakan tahapan krusial dalam *text mining* yang berfungsi untuk membersihkan dan menstandarisasi data tekstual mentah (*raw data*) sebelum diproses lebih lanjut oleh model komputasi. Data teks yang

diekstraksi, seperti lirik lagu, umumnya masih bersifat tidak terstruktur dan mengandung banyak gangguan (*noise*), seperti variasi huruf, tanda baca, serta kata-kata yang tidak memiliki nilai analitis (Manueke et al., 2023). Oleh karena itu, tahapan pra-pemrosesan diimplementasikan untuk mengubah teks bahasa alami tersebut menjadi format yang lebih terstruktur dan seragam, sehingga algoritma pembelajaran mesin (*machine learning*) dapat mengekstraksi fitur kata dengan tingkat akurasi yang lebih optimal (Sabda Husada & Yulianto, 2026).

### **2.7.1 Case Folding**

*Case folding* adalah proses penyeragaman seluruh karakter huruf pada korpus teks menjadi bentuk huruf kecil (*lowercase*). Proses ini bertujuan untuk menghindari inkonsistensi pembacaan matriks oleh sistem, karena secara komputasional, karakter huruf kapital dan huruf kecil sering kali direpresentasikan dengan nilai biner yang berbeda. Melalui implementasi *case folding*, kata yang letaknya di awal kalimat maupun di tengah kalimat akan dievaluasi sebagai satu entitas string yang sama dalam tahapan perhitungan bobot selanjutnya (Ramadhan As'ari et al., 2025).

### **2.7.2 Cleaning**

*Cleaning* atau pembersihan merupakan tahapan filtrasi untuk menghapus elemen-elemen teks yang tidak relevan dan tidak memberikan kontribusi makna terhadap proses analisis emosi. Pada tahapan ini, seluruh karakter non-alfabetik seperti angka, tanda baca, simbol (*special characters*), hingga elemen seperti tautan (*URL*) akan dieliminasi dari dokumen (Manueke et al., 2023). Pembersihan ini dilakukan untuk memastikan bahwa sistem komputasi hanya akan memproses susunan kata murni yang menyusun struktur lirik lagu tersebut.

### **2.7.3 Tokenizing**

*Tokenizing* merupakan tahapan pemecahan atau pemotongan serangkaian teks kalimat utuh menjadi unit-unit kata tunggal yang berdiri sendiri, yang dikenal sebagai token. Proses ini umumnya memisahkan *string* teks berdasarkan spasi atau karakter pemisah lainnya. Pemecahan kalimat menjadi token tunggal merupakan langkah esensial karena perhitungan probabilitas dan klasifikasi sentimen pada algoritma akan dieksekusi berdasarkan frekuensi kemunculan dari masing-masing unit token tersebut (Tri Widodo & Setyawan, 2025).

### **2.7.4 Stopword Removal**

*Stopword removal* adalah tahapan penyaringan untuk mengeliminasi kata-kata umum yang sering muncul di dalam teks namun tidak memiliki bobot makna spesifik untuk keperluan klasifikasi polaritas. Contoh kata yang masuk dalam daftar *stopword* meliputi kata hubung, kata ganti, atau kata depan, seperti "dan", "di", "ke", "yang", atau "dari". Dengan menghapus *stopword*, dimensi ruang data akan berkurang secara signifikan

sehingga proses komputasi mesin menjadi lebih efisien dan difokuskan secara khusus pada kata kunci yang mendefinisikan emosi (Sabda Husada & Yulianto, 2026).

### 2.7.5 Stemming

*Stemming* adalah proses analisis morfologis yang bertujuan untuk mengubah token kata yang memiliki imbuhan (prefiks, sufiks, infiks, maupun konfiks) kembali menjadi bentuk kata dasarnya (*root word*). Sebagai representasi, sistem akan mentransformasi variasi turunan kata menjadi entitas kata dasar tunggal. Tahapan ini beroperasi secara esensial dalam pemrosesan teks bahasa Indonesia guna mengurangi redundansi kamus data dan memastikan bahwa variasi leksikal yang berasal dari akar makna yang sama akan dikalkulasi sebagai satu fitur probabilitas yang ekuivalen (Ramadhan As'ari et al., 2025).

### 2.8 Term Frequency-Inverse Document Frequency (TF-IDF)

*Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan salah satu algoritma pembobotan kata yang paling fundamental dalam disiplin *text mining* dan pemrosesan bahasa alami. Metode ini digunakan untuk mengevaluasi dan mengkalkulasi seberapa penting sebuah kata (*term*) terhadap suatu dokumen tertentu di dalam sebuah kumpulan korpus teks (Virginia & Irsyad, 2025). Dalam proses komputasi, teks lirik lagu yang telah melewati tahapan pra-pemrosesan (*pre-processing*) akan diekstraksi nilai bobot numeriknya menggunakan algoritma ini, sehingga mesin dapat mengidentifikasi relevansi dari setiap kata sebelum diklasifikasikan ke dalam kelas sentimen (Syafaah & Lestandy, 2022).

Pendekatan komputasi pada metode ini mengintegrasikan dua parameter matematis utama. Parameter pertama adalah *Term Frequency* (TF), yang berfungsi untuk mengukur frekuensi kemunculan sebuah kata spesifik di dalam satu dokumen tunggal. Semakin sering kata tersebut muncul di dalam lirik, semakin tinggi nilai probabilitas lokalnya. Parameter kedua adalah *Inverse Document Frequency* (IDF), yang bertindak sebagai fungsi normalisasi untuk mengurangi atau memberi penalti pada bobot kata-kata umum yang muncul di hampir seluruh dokumen dalam korpus (Virginia & Irsyad, 2025). Melalui perpaduan parameter ini, TF-IDF mampu menyoroti kata kunci unik yang menjadi representasi utama dari dokumen tersebut. Secara matematis, kalkulasi nilai bobot akhir (*weight*) dari setiap kata diperoleh dengan mengalikan nilai frekuensi kemunculan (TF) dan nilai keunikan kata (IDF). Persamaan matematis pembobotan TF-IDF dijabarkan sebagai berikut:

$$W_{t,d} = TF(t,d) \times IDF(t)$$

Untuk memperoleh nilai probabilitas balikan dokumen (IDF), rumusan logaritmik yang digunakan adalah:

$$IDF(t) = \log\left(\frac{N}{df(t)}\right)$$

Keterangan

$W_{t,d}$  : Bobot akhir representasi kata (*term*)  $t$  dalam dokumen  $d$

$TF(t, d)$  : Bobot akhir representasi kata (*term*)  $t$  dalam dokumen  $d$

$N$  : Total keseluruhan jumlah dokumen di dalam korpus teks

$df(t)$  : Jumlah total dokumen yang mengandung kata (*term*)  $t$

## 2.9 Algoritma Naïve Bayes

Algoritma *Naive Bayes* merupakan salah satu metode klasifikasi probabilitas terawasi (*supervised learning*) yang berakar dari konsep statistika Teorema Bayes (Riyandona et al., 2025). Dalam bidang penambangan teks (*text mining*), algoritma ini sangat populer dan secara luas diimplementasikan untuk mengeksekusi proses analisis sentimen (Tri Widodo & Setyawan, 2025). Penggunaan *Naive Bayes* dinilai sangat relevan untuk mengklasifikasikan lirik lagu karena algoritma ini memiliki efisiensi komputasi yang tinggi dan kehandalan dalam memproses kumpulan data dengan dimensi atribut yang sangat besar (Ramadlan As'ari et al., 2025).

Pendekatan klasifikasi pada algoritma ini disebut "*naive*" atau lugu karena didasarkan pada asumsi penyederhanaan yang menganggap bahwa setiap atribut atau fitur kata di dalam sebuah dokumen bersifat independen yang kuat antarkondisinya (Ode et al., 2022). Asumsi independensi ini berarti bahwa probabilitas kemunculan sebuah kata di dalam teks dianggap tidak memiliki keterikatan struktural atau dampak langsung terhadap kemunculan kata lainnya. Meskipun kondisi kebebasan atribut tersebut jarang terjadi secara mutlak dalam tata bahasa alami, klasifikasi *Naive Bayes* terbukti mampu mengalkulasi dan memprediksi kelas polaritas emosi dengan tingkat akurasi yang tinggi (Josen Limbong et al., 2022).

Secara matematis, klasifikasi teks dieksekusi dengan mengalkulasi nilai probabilitas tertinggi dari suatu kelas target (misalnya sentimen positif, negatif, atau netral) berdasarkan keberadaan fitur-fitur kata pendukungnya. Persamaan probabilitas pada metode *Naive Bayes* dirumuskan melalui persamaan berikut:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

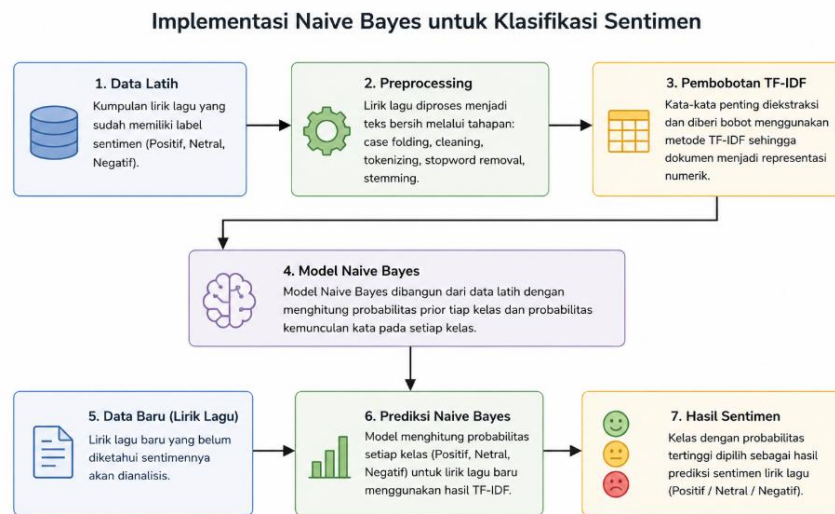
Keterangan:

$P(C|X)$  : Peluang bersyarat dari kelas sentimen  $C$  apabila diketahui dokumen  $X$  (*posterior probability*).

$P(X|C)$  : Peluang kemunculan fitur dokumen  $X$  pada kelas sentimen  $C$  (*likelihood*).

$P(C)$  : Peluang kemunculan kelas sentimen  $C$  secara global (*prior probability*).

$P(X)$  : Peluang kemunculan dokumen  $X$  secara global (*evidence*).



**Gambar 2. 1 Alur Implementasi Metode *Multinomial Naive Bayes***

## 2.10 Evaluasi Model

Tahapan evaluasi model merupakan fase krusial dalam siklus komputasi yang difungsikan untuk mengukur tingkat kinerja, kehandalan, serta validitas dari algoritma pembelajaran mesin (*machine learning*) yang telah diimplementasikan (Hapsari & Pandya Wisesa, 2025). Pada penelitian berbasis klasifikasi teks, pengujian ini bertujuan untuk memvalidasi seberapa akurat sistem dalam memprediksi kelas polaritas emosi jika dikomparasikan dengan label aktual atau data sesungguhnya (Riyandona et al., 2025). Hasil dari evaluasi ini akan merepresentasikan kelayakan sistem komputasi dalam mengekstraksi dan mengklasifikasikan lirik lagu secara objektif.

### 2.10.1 *Confusion Matrix*

*Confusion matrix* atau matriks kebingungan adalah alat pengujian performa berupa representasi tabel yang memvisualisasikan matriks perbandingan antara hasil prediksi algoritma komputasi dengan data target atau label kelas aktual (Riyandona et al., 2025). Tabel ini memberikan rincian komprehensif mengenai tingkat keberhasilan maupun kesalahan sistem dalam mengklasifikasikan teks. Matriks ini mengekstraksi empat parameter dasar pengujian, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Keempat parameter inilah yang nantinya dikalkulasikan untuk menghasilkan metrik evaluasi kuantitatif secara lebih spesifik.

### 2.10.2 *Accuracy*

*Accuracy* atau akurasi adalah metrik pengukuran yang merepresentasikan persentase ketepatan atau kecocokan keseluruhan hasil prediksi model algoritma terhadap total keseluruhan data teks yang diuji

(Hapsari & Pandya Wisesa, 2025). Metrik ini menjadi indikator umum yang paling sering diandalkan untuk melihat seberapa baik sistem membedakan setiap kelas sentimen secara global. Secara matematis, tingkat akurasi dirumuskan melalui persamaan berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Keterangan:

**TP** (*True Positive*) : Jumlah data polaritas positif yang berhasil diprediksi dengan benar sebagai positif oleh sistem.

**TN** (*True Negative*) : Jumlah data polaritas negatif yang berhasil diprediksi dengan benar sebagai negatif oleh sistem.

**FP** (*False Positive*) : Jumlah data polaritas negatif yang salah diprediksi sebagai positif oleh sistem.

**FN** (*False Negative*): Jumlah data polaritas positif yang salah diprediksi sebagai negatif oleh sistem.

### 2.10.3 Precision

*Precision* atau presisi merupakan metrik spesifik yang mengukur tingkat ketepatan prediksi positif yang dihasilkan oleh sistem (Riyandona et al., 2025). Metrik ini menunjukkan rasio antara jumlah data yang diprediksi benar positif oleh model dengan total seluruh data yang diprediksi sebagai positif (baik yang benar maupun yang salah prediksi). Persamaan matematis untuk kalkulasi *precision* adalah sebagai berikut:

$$Precision = \frac{TP}{TP + FP}$$

### 2.10.4 Recall

*Recall* atau tingkat kelengkapan adalah metrik komputasional yang mengalkulasi rasio jumlah prediksi benar positif dibandingkan dengan total keseluruhan data aktual yang bernilai positif di dalam korpus uji (Hapsari & Pandya Wisesa, 2025). Metrik ini sangat esensial untuk mengetahui sejauh mana algoritma mampu mengenali dan memanggil kembali (*recall*) seluruh informasi yang relevan di dalam sebuah basis data tekstual. Rumusan matematis *recall* dijabarkan sebagai berikut:

$$Recall = \frac{TP}{TP + FN}$$

### 2.10.5 F1-Score

*F1-Score* merupakan metrik evaluasi yang mempresentasikan nilai rata-rata harmonik dari *precision* dan *recall* (Hapsari & Pandya Wisesa, 2025). Penggunaan metrik ini sangat relevan apabila terdapat distribusi kelas sampel yang tidak seimbang (*imbalanced data*) di dalam *dataset*.

Melalui kalkulasi *F1-Score*, peneliti dapat memperoleh representasi tunggal yang menyeimbangkan *trade-off* antara nilai presisi dan tingkat kelengkapan pemanggilan data. Persamaan kalkulasi *F1-Score* dirumuskan melalui persamaan berikut:

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Keterangan:

**Precision** : Tingkat rasio ketepatan prediksi positif dari sistem.

**Recall** : Tingkat kelengkapan sistem dalam memprediksi data positif secara aktual.

## 2.11 Studi Literatur

Studi literatur berfungsi untuk mengkaji penelitian terdahulu guna membangun landasan teoretis, mengevaluasi metodologi komputasi, serta mengidentifikasi celah penelitian (*research gap*) dan nilai kebaruan (*novelty*). Pada penelitian ini, rujukan literatur dikerucutkan pada sepuluh jurnal utama yang berfokus pada implementasi *Natural Language Processing* (NLP), penambangan teks (*text mining*), algoritma *Naive Bayes*, serta pemrosesan teks pada lirik lagu. Komparasi dari kesepuluh literatur terdahulu tersebut dijabarkan secara sistematis pada Tabel 2.1 berikut.

**Tabel 2. 1 Studi Literatur**

| No | Peneliti & Tahun               | Judul Penelitian                                                                            | Metode/Algoritma                               | Hasil Penelitian                                                                                                                           | Kontribusi terhadap Riset                                                                                                    |
|----|--------------------------------|---------------------------------------------------------------------------------------------|------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| 1  | Manueke et al. (2023)          | Analisis Sentimen Lirik Lagu Populer Dengan Pemanfaatan Teknik Natural Language Processing. | Natural Language Processing (NLP), Text Mining | Implementasi NLP terbukti efektif dalam mentransformasi teks lirik lagu yang ambigu menjadi data yang dapat diekstraksi nilai sentimennya. | Menjadi rujukan landasan teori bahwa NLP merupakan metode yang paling tepat untuk memproses struktur bahasa pada lirik lagu. |
| 2  | Sabda Husada & Yulianto (2026) | Analisis Emosi dalam Lirik Lagu menggunakan Natural Language Processing.                    | Natural Language Processing (NLP)              | Sistem komputasi berhasil mengenali dan mengklasifikasikan ungkapan emosi spesifik dari berbagai                                           | Memberikan landasan observasi mengenai pemetaan kelas polaritas emosi pada teks sastra kontemporer seperti lirik lagu.       |

|   |                                  |                                                                                              |                                     |                                                                                                                                     |                                                                                                                                     |
|---|----------------------------------|----------------------------------------------------------------------------------------------|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
|   |                                  |                                                                                              |                                     | objek teks karya musik.                                                                                                             |                                                                                                                                     |
| 3 | Ode et al. (2022)                | Klasifikasi Lirik Lagu Bertema Lingkungan dengan Metode Naive Bayes.                         | Naive Bayes Classifier              | Algoritma Naive Bayes menghasilkan tingkat akurasi yang tinggi saat diterapkan pada korpus dokumen lirik bertema khusus.            | Memvalidasi kehandalan algoritma Naive Bayes dalam memproses sekumpulan data teks lirik yang memiliki dimensi atribut sangat besar. |
| 4 | Kholida Zia Abidin et al. (2023) | Systematic Literature Review terhadap Klasifikasi Emosi pada Lirik Lagu.                     | Systematic Literature Review (SLR)  | Pemetaan berbagai variasi algoritma pembelajaran mesin terbukti sangat krusial dalam menganalisis ungkapan afektif lirik lagu.      | Menjustifikasi urgensi penelitian analisis emosi pada karya musik dan membantu menentukan posisi kebaruan (novelty) penelitian ini. |
| 5 | Riyandona et al. (2025)          | Implementasi Model Analisis Sentimen Terhadap Grup Musik BTS Menggunakan Metode Naive Bayes. | Naive Bayes, Analisis Sentimen      | Model probabilitas bersyarat pada Naive Bayes terbukti sangat efisien dalam memproses sentimen opini publik terkait industri musik. | Memperkuat argumen pemilihan algoritma probabilitas karena ketepatan metrik evaluasinya pada objek bertema musik.                   |
| 6 | Josen Limbong et al. (2022)      | Analisis Klasifikasi Sentimen Ulasan Berbasis Word Cloud dengan Metode Naive Bayes.          | Naive Bayes, Visualisasi Word Cloud | Penggabungan hasil klasifikasi dengan awan kata (word cloud) sangat efektif memetakan probabilitas kata dominan dari teks.          | Menjadi referensi utama untuk perancangan fitur ekstraksi topik dan visualisasi kata kunci (term) pada antarmuka sistem.            |
| 7 | Odhianto et al. (2025)           | Optimalisasi Latent                                                                          | Text Mining,                        | Ekstraksi topik secara akurat                                                                                                       | Memberikan referensi mengenai                                                                                                       |

|    |                                    |                                                                                       |                                      |                                                                                                                                     |                                                                                                                                    |
|----|------------------------------------|---------------------------------------------------------------------------------------|--------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
|    |                                    | Dirichlet Allocation untuk Ekstraksi Topik Utama dalam Teks Dongeng.                  | Ekstraksi Topik                      | berhasil menemukan tema naratif tersembunyi pada dokumen sastra yang tidak terstruktur.                                             | pentingnya ekstraksi kata kunci untuk merangkum tema utama pada karya berbasis teks puitis/sastra.                                 |
| 8  | Virginia & Irsyad (2025)           | Rekomendasi Lagu Berdasarkan Jenis Mood Menggunakan TF-IDF.                           | Pembobotan TF-IDF, Cosine Similarity | Parameter bobot TF-IDF mampu memetakan frekuensi kata yang paling relevan dengan suasana hati (mood) dari sebuah lagu.              | Menjadi landasan matematis penerapan algoritma pembobotan TF-IDF untuk menyeleksi fitur kata pada lirik R&B.                       |
| 9  | Salsabila Nuzula & Djauhari (2025) | Discourse of Apology and Plea in R&B Lyrics: A Pragmatic Analysis.                    | Analisis Pragmatis, Linguistik       | Secara leksikal, lirik lagu R&B terbukti memiliki kedalaman wacana emosional tinggi, sarat akan diksi permohonan dan introspeksi.   | Memberikan landasan karakteristik linguistik dan justifikasi kuat mengapa lirik R&B dan neo-soul dipilih sebagai objek penelitian. |
| 10 | Ramadhan As'ari et al. (2025)      | Analisis Sentimen Ulasan Pengguna Aplikasi Spotify Menggunakan Algoritma Naive Bayes. | Naive Bayes, Pre-processing Teks     | Tahapan pra-pemrosesan yang dipadukan dengan Naive Bayes menghasilkan evaluasi matriks kebingungan (confusion matrix) yang optimal. | Sebagai acuan tahapan sistematis pra-pemrosesan teks (pre-processing) untuk membersihkan gangguan (noise) pada data linguistik.    |

## BAB III METODOLOGI PENELITIAN

### 3.1 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental yang berfokus pada penerapan komputasi berbasis teks atau *text mining*. Pendekatan kuantitatif dipilih karena analisis variasi emosi dan klasifikasi polaritas dalam penelitian ini diukur serta dieksekusi menggunakan pemodelan statistik berbasis angka, yang dihasilkan dari perhitungan matematis pembobotan tekstual lirik lagu. Metode utama yang diterapkan dalam penelitian ini adalah penambangan teks (*text mining*), sebuah disiplin turunan dari *data mining* yang secara spesifik mengekstraksi pola informasi berharga, pengetahuan baru, atau karakteristik tersembunyi dari sekumpulan data tekstual yang bersifat tidak terstruktur (*unstructured data*).

#### 3.1.1 Kerangka Penelitian

Guna memastikan proses penambangan teks berjalan secara sistematis, penelitian ini mengadopsi kerangka kerja *Knowledge Discovery in Databases* (KDD) (Syafaah & Lestandy, 2022). Kerangka kerja *Knowledge Discovery in Databases* (KDD) dipilih karena memiliki tahapan terstruktur yang sangat adaptif dalam mengolah data dari bentuk mentah hingga menjadi pengetahuan (*knowledge*) yang siap diinterpretasikan. Secara umum, alur kronologis penelitian komputasi lirik lagu ini berjalan melalui tahapan-tahapan makro sebagai berikut:

1. **Fase Seleksi Data (*Data Selection*):** Melakukan pengumpulan data lirik lagu R&B dan *neo-soul* dari platform digital, kemudian menyaring dokumen tersebut berdasarkan parameter inklusi untuk membentuk korpus *dataset* yang valid dan memenuhi standar minimal kuantitas analisis.
2. **Fase Pra-Pemrosesan (*Pre-processing*):** Mentransformasikan data lirik mentah yang sarat akan gangguan (*noise*) ke dalam bentuk teks yang bersih dan seragam melalui teknik sekuensial *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*.
3. **Fase Transformasi (*Transformation*):** Mengonversi kata dasar hasil pembersihan menjadi matriks koordinat vektor numerik menggunakan metode pembobotan frekuensi dan keunikan kata melalui algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF).
4. **Fase Penambangan Teks (*Text Mining*):** Mengimplementasikan algoritma klasifikasi berbasis probabilitas *Naive Bayes* pada matriks numerik untuk menentukan label polaritas sentimen pada dokumen lirik, serta melakukan identifikasi topik berdasarkan kemunculan kata-kata dominan yang merepresentasikan tema utama pada korpus lirik.
5. **Fase Evaluasi (*Evaluation*):** Mengukur performa model klasifikasi menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengetahui tingkat keberhasilan dan kehandalan algoritma dalam melakukan klasifikasi sentimen.

6. **Fase Perancangan Antarmuka Sistem:** Merancang struktur menu dan tampilan layar aplikasi sebagai media interaksi pengguna dalam melakukan pengolahan data lirik, analisis sentimen, serta visualisasi hasil ekstraksi topik.

### 3.1.2 Arsitektur Sistem

Sistem klasifikasi sentimen dan identifikasi topik dominan ini dibangun menggunakan arsitektur gabungan (*hybrid*) berbasis web. Pengembangan antarmuka pengguna (*Graphical User Interface*) dan pengelolaan struktur basis data relasional (MySQL) diimplementasikan menggunakan bahasa pemrograman PHP yang bertindak sebagai gerbang interaksi utama sistem. Di sisi lain, keseluruhan komputasi kecerdasan buatan—yang mencakup tahapan pra-pemrosesan teks, pembobotan TF-IDF, ekstraksi fitur dominan, hingga pelokasian model Multinomial Naive Bayes—dieksekusi secara terpusat oleh mesin *Machine Learning* menggunakan bahasa pemrograman Python. Kedua lingkungan pengembangan perangkat lunak ini diintegrasikan secara dinamis melalui fungsi jembatan *shell execution* (*shell\_exec*), sehingga sistem mampu membaca pembaruan korpus lirik, melatih ulang model probabilitas, serta memprediksi sentimen data secara *real-time* langsung melalui peramban (*browser*).

## 3.2 Dataset dan Sumber Data Lirik Lagu

Penelitian ini menggunakan *dataset* sekunder berupa sekumpulan data teks lirik lagu kontemporer yang secara spesifik diklasifikasikan ke dalam genre R&B dan *neo-soul*. Proses pengumpulan data lirik utama dilakukan melalui kombinasi pengumpulan manual dan pemanfaatan sumber data lirik digital dari platform pangkalan data lirik musik terbesar, yakni Genius.com. Pemilihan Genius.com sebagai sumber data didasarkan pada tingkat akurasi liriknya yang divalidasi langsung oleh komunitas dan musisi, sehingga sangat meminimalisir kesalahan penulisan teks (*typo*) pada korpus.

Di samping penarikan data secara otomatis, pengumpulan data juga dilakukan secara manual untuk mengakomodasi karya-karya dari ekosistem musik independen lokal yang relevan dengan genre tersebut. Total keseluruhan data yang berhasil dihimpun dan digunakan dalam penelitian ini adalah sebanyak **207 baris dokumen lirik lagu**. Angka tersebut telah dipastikan memenuhi ambang batas minimal kelayakan pemrosesan *text mining* sesuai dengan standar pedoman akademik.

*Dataset* yang telah dikurasi tersebut kemudian disimpan ke dalam format *Comma Separated Values* (.csv) agar kompatibel dan mudah diimpor melalui fitur *browse* pada antarmuka *Graphical User Interface* (GUI) sistem yang dirancang. Setiap baris data pada korpus ini memuat tiga atribut utama yang menjadi parameter analisis, yaitu nama artis, judul lagu, dan bait lirik mentah (*raw text*) yang memuat keseluruhan isi lagu.

Tabel 3.1 di bawah ini menampilkan cuplikan representatif dari *dataset* lirik lagu yang digunakan dalam sistem, mencakup rentang karya

dari musisi arus utama internasional hingga salah satu sampel karya independen lokal yang merepresentasikan karakteristik narasi *neo-soul* secara personal:

**Tabel 3. 1 Sampel Dataset Lirik Lagu**

| ID | Artis                | Judul Lagu         | Lirik Lagu (Data Teks Mentah)                                           |
|----|----------------------|--------------------|-------------------------------------------------------------------------|
| 1  | Mac Ayres            | Easy               | It is easy to fall in love with you but hard to stay...                 |
| 2  | xdika                | Teduh              | Langkahmu lirik namun, menyibak gelap tanpa ragu...                     |
| 3  | Daniel Caesar        | Get You            | Through drought and famine natural disasters my baby has been around... |
| 4  | Glenn Fredly         | Sedih Tak Berujung | Saat menjelang hari-hari bahagiamu aku memilih diam dalam sepi...       |
| 5  | Maliq & D'Essentials | Untitled           | Tidakkah cukup yang engkau lihat pertemanan ini sendiri...              |

### 3.3 Seleksi Data (*Data Selection*)

Tahap seleksi data (*data selection*) merupakan langkah prosedural pertama dan fundamental untuk membangun fondasi informasi sebelum masuk ke tahapan komputasi bahasa alami. Pada fase ini, dilakukan pengumpulan sekaligus penyaringan (*filtering*) terhadap sekumpulan data teks mentah (*raw data*) yang bersumber dari platform Genius.com, yang dipadukan dengan hasil kurasi lirik secara manual. Penggunaan sumber data yang komprehensif ini bertujuan untuk memperoleh variasi leksikal yang kaya, sekaligus mengakomodasi karya-karya dari musisi arus utama (*mainstream*) maupun karya dari ekosistem musik independen yang bergerak secara spesifik di jalur R&B dan *neo-soul*.

Mengingat volume data tekstual di platform digital sangat masif dan bersifat heterogen, proses seleksi diterapkan menggunakan parameter inklusi dan eksklusi yang ketat. Proses utama difokuskan pada isolasi genre, di mana sistem hanya mengekstrak lirik yang secara eksplisit dikategorikan ke dalam *metadata* genre R&B dan *neo-soul* kontemporer. Selanjutnya, dilakukan penyaringan anomali data untuk mengeliminasi lagu-lagu instrumental (tanpa lirik), versi pertunjukan langsung (*live version*), maupun versi paduan (*remix*) yang berpotensi menghasilkan duplikasi kalimat atau pergeseran topik. Teks lirik yang terlalu pendek atau tidak memiliki struktur bait yang utuh juga dibuang dari korpus data untuk mencegah terjadinya bias (*noise*) pada saat algoritma melakukan pembobotan kata di tahap selanjutnya.

Selain penyaringan kualitas tekstual, tahapan ini juga mencakup proses penghapusan data ganda (*data deduplication*) untuk memastikan bahwa setiap judul lagu yang masuk ke dalam *dataset* merupakan entitas analisis yang unik. Proses seleksi ini dikurasi secara ketat guna memverifikasi bahwa *dataset* akhir telah memenuhi dan melampaui ambang batas minimal kelayakan pemrosesan

*text mining* sesuai dengan standar kewajiban akademik, yakni mencapai hasil akhir sebanyak **207 baris** dokumen representatif.

Korpus data lirik pilihan yang telah terverifikasi tersebut kemudian diseragamkan struktur atributnya—yang mencakup nama artis, judul lagu, dan teks lirik—untuk selanjutnya disimpan secara terpusat ke dalam format berkas *Comma Separated Values* (.csv). Format ini dipastikan kompatibel untuk dieksekusi pada tahapan pra-pemrosesan (*pre-processing*).

### 3.4 Pra-Pemrosesan (*Pre-Processing*)

Tahap Pra-pemrosesan data merupakan tahapan krusial dalam komputasi *text mining* yang bertujuan untuk membersihkan dan menormalisasi *dataset* teks mentah sebelum masuk ke tahap ekstraksi fitur. Karena dokumen lirik lagu umumnya mengandung struktur teks yang tidak baku, karakter spesial, serta banyak kosakata yang tidak memiliki bobot sentimen, tahap pra-pemrosesan mutlak diperlukan untuk mencegah penurunan tingkat akurasi pada algoritma Multinomial Naive Bayes. Dalam penelitian ini, proses pra-pemrosesan diimplementasikan menggunakan bahasa pemrograman Python dengan memanfaatkan pustaka pemrosesan bahasa alami *Natural Language Toolkit* (NLTK) dan Sastrawi melalui lima tahapan sekuensial berikut:

#### 3.4.1 *Case Folding*

Tahap *case folding* adalah proses penyeragaman seluruh karakter huruf di dalam dokumen teks menjadi huruf kecil (*lowercase*). Pada tahap ini, mesin mengonversi seluruh huruf kapital yang terdapat pada awal kalimat, nama, maupun penekanan lirik menjadi huruf kecil. Hal ini bertujuan agar algoritma komputasi menganggap kata yang sama dengan penulisan kapitalisasi yang berbeda (contoh: "*Love*", "*LOVE*", dan "*love*") sebagai satu entitas kata yang identik.

**Tabel 3. 2 *Case Folding***

| Sebelum <i>Case Folding</i>                                                  | Sesudah <i>Case Folding</i>                                                  |
|------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| Saat menjelang hari-hari bahagiamu, aku memilih diam dalam sepi... Oh, yeah! | saat menjelang hari-hari bahagiamu, aku memilih diam dalam sepi... oh, yeah! |

#### 3.4.2 *Pembersihan (Cleaning)*

Tahap *cleaning* merupakan proses pembersihan dokumen dari karakter-karakter yang tidak memiliki nilai analitik. Menggunakan metode *Regular Expression* (Regex) pada Python, sistem secara otomatis menghapus seluruh angka, tanda baca, simbol, serta karakter non-alfabet lainnya. Pada konteks *dataset* lirik lagu, tahap ini secara spesifik sangat berguna untuk menghilangkan tanda kurung siku atau kurung lengkung yang biasanya memuat penanda struktur

lagu seperti [*Chorus*], [*Verse*], maupun (*Intro*) yang sering terbawa saat proses penarikan data ( *scraping* ).

**Tabel 3. 3 Cleaning**

| Sebelum <i>Cleaning</i>                                                            | Sesudah <i>Cleaning</i>                                                      |
|------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| saat menjelang hari-hari<br>bahagiamu, aku memilih diam<br>dalam sepi... oh, yeah! | saat menjelang hari hari<br>bahagiamu aku memilih diam<br>dalam sepi oh yeah |

### 3.4.3 Tokenizing

Setelah teks lirik bersih dari karakter non-alfabet dan seragam dalam format huruf kecil, langkah selanjutnya adalah *tokenizing*. Tahap ini memecah untaian kalimat panjang di dalam lirik menjadi potongan-potongan kata tunggal atau *token* yang berdiri sendiri. Pemisahan ini umumnya menggunakan spasi sebagai karakter pembatas (*delimiter*), sehingga mesin dapat menghitung frekuensi dan menganalisis probabilitas masing-masing kata secara individual pada tahapan pemodelan.

**Tabel 3. 4 Tokenizing**

| Sebelum <i>Tokenizing</i>                                                    | Sesudah <i>Tokenizing</i>                                                                                            |
|------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|
| saat menjelang hari hari<br>bahagiamu aku memilih diam<br>dalam sepi oh yeah | ["saat", "menjelang", "hari",<br>"hari", "bahagiamu", "aku",<br>"memilih", "diam", "dalam",<br>"sepi", "oh", "yeah"] |

### 3.4.4 Penghapusan Kata Umum (*Stopword Removal*)

*Stopword removal* adalah tahap penyaringan untuk membuang kata-kata umum yang sering muncul namun tidak membawa informasi atau makna sentimen yang signifikan (seperti kata hubung, kata ganti, dan kata depan). Sistem ini menerapkan penyaringan berlapis menggunakan kamus bawaan dari NLTK untuk bahasa Inggris dan Sastrawi untuk bahasa Indonesia.

Sebagai bentuk optimasi pada penelitian ini, ditambahkan pula sebuah kamus *custom stopwords* yang dirancang khusus untuk mengeliminasi *noise* leksikal pada genre musik R&B dan *neo-soul*. Kata-kata dominan yang bersifat gumaman vokal, *ad-libs*, atau ekspresi musikal yang tidak memiliki muatan sentimen (seperti *yeah*, *ooh*, *mmm*, *uh*, *ah*) secara otomatis dihapus dari korpus. Pembuangan *noise* spesifik ini terbukti krusial untuk mencegah algoritma mengalami kebingungan (*overlapping features*) dalam membedakan kelas sentimen positif dan negatif.

**Tabel 3. 5 Stopword Removal**

| Sebelum <i>Stopword Removal</i>                                                                             | Sesudah <i>Stopword Removal</i>                                       |
|-------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| ["saat", "menjelang", "hari", "hari", "bahagiamu", "aku", "memilih", "diam", "dalam", "sepi", "oh", "yeah"] | ["menjelang", "hari", "hari", "bahagiamu", "memilih", "diam", "sepi"] |

### 3.4.5 Stemming / Lemmatization

Tahap terakhir dari proses pra-pemrosesan adalah *stemming*, yaitu proses pemotongan atau transformasi *token* kata yang memiliki imbuhan (afiksasi) kembali ke bentuk kata dasarnya (*root word*). Proses ini dilakukan menggunakan algoritma aturan morfologi dari pustaka Sastrawi. Sebagai contoh, kata-kata yang memiliki variasi imbuhan akan diseragamkan menjadi satu bentuk dasar, sehingga mengurangi keragaman kosakata (*dimensionality reduction*) tanpa menghilangkan maknanya. Hasil akhir dari tahap *stemming* ini adalah *dataset* lirik yang telah bersih, padat, dan siap untuk dikonversi menjadi matriks angka melalui pembobotan TF-IDF.

**Tabel 3. 6 Stemming**

| Sebelum <i>Stemming</i>                                               | Sesudah <i>Stemming</i>                                        |
|-----------------------------------------------------------------------|----------------------------------------------------------------|
| ["menjelang", "hari", "hari", "bahagiamu", "memilih", "diam", "sepi"] | ["jelang", "hari", "hari", "bahagia", "pilih", "diam", "sepi"] |

### 3.5 Ekstraksi Fitur Menggunakan TF-IDF

Pada tahap ini, korpus teks lirik yang telah melewati seluruh proses pembersihan akan dikonversi menjadi representasi numerik menggunakan algoritma pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF). Proses ekstraksi fitur ini sangat krusial karena mesin *Machine Learning* tidak dapat memproses data dalam bentuk karakter alfabet murni, melainkan harus diubah ke dalam bentuk matriks vektor angka (Josen Limbong et al., 2022).

Secara matematis, TF-IDF mengevaluasi seberapa relevan dan penting sebuah kata di dalam satu lirik lagu relatif terhadap seluruh kumpulan lagu (*dataset*). *Term Frequency* (TF) bertugas menghitung jumlah frekuensi kemunculan sebuah kata di dalam satu lirik. Sementara itu, *Inverse Document Frequency* (IDF) bertindak sebagai penyeimbang dengan cara memberikan "penalti" atau penurunan bobot pada kata-kata umum yang muncul hampir di semua lagu, sehingga kata-kata spesifik yang unik akan mendapatkan bobot numerik yang lebih tinggi (Josen Limbong et al., 2022).

Lebih lanjut, sebagai bentuk ekstraksi informasi lanjutan dari penelitian ini, perhitungan frekuensi kata (*Term Frequency*) tidak hanya digunakan sebagai masukan untuk model klasifikasi, melainkan juga dimanfaatkan untuk melakukan identifikasi topik dominan. Berdasarkan

hasil pemeringkatan matriks TF-IDF, kosakata yang memiliki kemunculan tertinggi (seperti kata *love*, *baby*, *feel*, *time*, *life*) akan diekstraksi menjadi representasi tema utama dari korpus lirik. Pemunculan kata-kata dominan yang berkaitan dengan hubungan interpersonal ini membuktikan bahwa ekstraksi topik pada lirik lagu bergenre R&B dan *neo-soul* secara valid merefleksikan narasi romansa dan gejolak emosional (Salsabila Nuzula & Djauhari, 2025).

### 3.6 Klasifikasi Sentimen Menggunakan Multinomial Naive Bayes

Setelah dokumen teks berhasil direpresentasikan dalam bentuk matriks bobot numerik melalui TF-IDF, tahap komputasi inti selanjutnya adalah melakukan prediksi polaritas sentimen menggunakan algoritma probabilitas *Multinomial Naive Bayes* (MNB). Pemilihan varian *multinomial* dalam arsitektur sistem ini didasarkan pada kecocokannya secara matematis untuk menangani fitur data diskrit berupa perhitungan frekuensi, sehingga sangat ideal untuk disandingkan dengan matriks keluaran dari TF-IDF (Riyandona et al., 2025).

Secara konseptual, algoritma *Naive Bayes* bekerja berlandaskan pada Teorema Bayes dengan mengedepankan asumsi "naif" (*naive*), yakni menganggap bahwa kemunculan setiap fitur kata di dalam lirik lagu saling independen dan tidak memiliki keterkaitan dengan keberadaan kata lainnya (Riyandona et al., 2025). Sistem akan mengalkulasi probabilitas kemunculan matriks kata dari lirik yang diuji terhadap data latih (*training data*) yang telah diberi label (positif, netral, atau negatif) sebelumnya.

Dokumen lirik lagu baru yang dimasukkan ke dalam sistem kemudian akan diklasifikasikan ke dalam kelas sentimen yang mengantongi nilai probabilitas *posterior* (probabilitas akhir) tertinggi. Penggunaan *Multinomial Naive Bayes* diterapkan secara spesifik pada penelitian ini mengingat komputasinya yang sangat efisien dalam memori, memiliki waktu komputasi yang cepat, serta terbukti memiliki performa akurasi yang tangguh dalam menyelesaikan masalah klasifikasi pada kumpulan data teks (*text mining*) berdimensi tinggi, khususnya dalam analisis emosi lirik lagu (Sabda Husada & Yulianto, 2026).

### 3.7 Rancangan Basis Data

Rancangan basis data merupakan tahapan perancangan arsitektur penyimpanan data yang digunakan untuk menampung seluruh informasi sistem. Sistem klasifikasi sentimen ini menggunakan sistem manajemen basis data relasional MySQL. Mengingat sistem ini berfokus pada pemrosesan teks (*text mining*), arsitektur basis data dirancang secara efisien menggunakan dua entitas tabel utama, yaitu tabel admin untuk autentikasi keamanan dan tabel *dataset* untuk menyimpan korpus lirik beserta hasil transformasinya.

#### 3.7.1 Logical Record Structure (LRS)

Representasi struktur data secara logis tanpa ketergantungan pada mekanisme penyimpanan fisik didefinisikan melalui konsep *Logical Record Structure* (LRS). Model ini difungsikan untuk memvisualisasikan skema relasional antar-entitas, di mana integritas referensial data dijaga

melalui mekanisme pengaitan antara kunci utama (*primary key*) dan kunci tamu (*foreign key*). Pemahaman mengenai struktur ini penting untuk memastikan bahwa setiap elemen data dapat diakses dan dikelola secara efisien dalam sistem basis data relasional.

Dalam konteks sistem yang dibangun, implementasi LRS dirancang untuk menjamin kontinuitas alur informasi yang terintegrasi di dalam basis data tugasakhir\_andika, yang difokuskan pada relasi antara entitas *tbl\_admin* dan *tbl\_dataset*. Konektivitas antar-tabel dibangun menggunakan atribut identitas unik (*id\_admin*) yang berperan sebagai jembatan penghubung antara data autentikasi pengguna dengan pengelolaan korpus lirik lagu beserta seluruh riwayat transformasi teksnya. Konfigurasi lengkap mengenai struktur logika relasional tersebut disajikan pada Gambar 3.1.



**Gambar 3.1 Logical Record Structure**

### 3.7.2 Class Diagram

Untuk memodelkan struktur statis dari sistem basis data yang dibangun, digunakan *Class Diagram* yang merupakan bagian dari standar *Unified Modeling Language* (UML). Diagram ini memvisualisasikan skema sistem dengan memetakan kelas-kelas entitas, atribut yang dimilikinya, serta relasi antar kelas tersebut. Dalam konteks penelitian analisis sentimen lirik ini, *Class Diagram* berfungsi menggambarkan alur penyimpanan data, mulai dari tahap autentikasi hingga pencatatan hasil klasifikasi akhir. Struktur rinci dari diagram tersebut direpresentasikan pada Gambar 3.2.



**Gambar 3. 2 Class Diagram**

Berdasarkan rancangan visual di atas, sistem basis data dikonstruksi atas dua entitas utama yang saling terintegrasi. Deskripsi fungsional dari masing-masing entitas adalah sebagai berikut:

a. **tbl\_admin**

Entitas ini berperan sebagai repositori utama untuk menampung data kredensial pengguna yang memiliki hak akses penuh ke dalam sistem. Atribut kunci yang disimpan meliputi *id\_admin* bertipe *integer* sebagai penanda unik, serta *username* dan *password* bertipe *varchar* untuk proses autentikasi keamanan. Kelas ini memegang relasi *one-to-many* (1:M) terhadap entitas dataset, di mana satu admin memiliki wewenang untuk mengelola dan memproses banyak dokumen lirik lagu.

b. **tbl\_dataset**

Tabel ini difungsikan sebagai repositori utama atau basis pengetahuan (*knowledge base*) yang memuat rekapitulasi korpus data lirik beserta seluruh jejak rekam pra-pemrosesan teksnya. Informasi yang disimpan mencakup *id* sebagai identitas unik, *id\_admin* sebagai atribut penghubung, serta *metadata* musik berupa *artist*, *title*, dan *genre*. Lebih lanjut, entitas ini mendokumentasikan setiap tahapan transformasi *text mining*, mulai dari lirik mentah (*lyrics*), luaran *case\_folding*, *cleaning*, *tokenizing*, *stopword\_removed*, hingga *stemmed* dan *clean\_lyrics*. Selain itu, kronologi penyimpanan dicatat dalam atribut *upload\_date*, sedangkan parameter target dan prediksi dari algoritma klasifikasi direkam secara terstruktur pada atribut *sentiment\_label* dan *sentiment\_result*.

### 3.7.3 Spesifikasi Basis Data

Spesifikasi basis data menjabarkan rincian tipe data, panjang karakter, dan fungsi dari masing-masing atribut (kolom) yang terdapat di

dalam tabel basis data. Berikut adalah spesifikasi dari kedua tabel yang digunakan dalam sistem:

a. Tabel *tbl\_admin*

Tabel *tbl\_admin* berfungsi sebagai repositori sentral untuk autentikasi dalam arsitektur sistem ini. Tabel ini didesain khusus untuk menampung seluruh himpunan data pengguna yang memiliki hak akses (*privilege*) sebagai pengelola sistem. Identitas dan karakteristik teknis dari tabel ini adalah sebagai berikut:

Nama Database : tugasakhir\_andika

Nama Table : *tbl\_admin*

Primary Key : *id\_admin*

Isi : Data Autentikasi Pengelola Sistem

Untuk memberikan gambaran yang lebih komprehensif mengenai struktur penyimpanan, rincian atribut kolom yang mencakup nama *field*, tipe data, kapasitas penyimpanan (lebar), serta deskripsi fungsional masing-masing atribut disajikan dalam Tabel 3.7.

**Tabel 3. 7 Struktur Tabel *tbl\_admin***

| Nama Field      | Tipe Data | Panjang | Keterangan                      |
|-----------------|-----------|---------|---------------------------------|
| <i>id_admin</i> | int       | 11      | ID unik pengelola sistem (PK)   |
| <i>username</i> | varchar   | 50      | Nama pengguna untuk autentikasi |
| <i>password</i> | varchar   | 255     | Kata sandi pengguna             |

b. Tabel *tbl\_dataset*

Tabel *tbl\_dataset* dikonstruksi sebagai entitas referensi utama atau wadah komputasi teks (*text mining*) dalam kerangka penelitian ini. Fungsi fundamental dari tabel ini adalah untuk mengarsipkan data lirik lagu mentah beserta rekam jejak hasil pra-pemrosesan dan label sentimennya. Adapun spesifikasi teknis yang diterapkan pada tabel ini adalah sebagai berikut:

Nama Database : tugasakhir\_andika

Nama Table : *tbl\_dataset*

Primary Key : *id*

Foreign Key : *id\_admin*

Isi : Data Lirik Lagu dan Hasil *Text Mining*

Guna memberikan pemahaman mendalam mengenai atribut yang digunakan dalam proses komputasi, rincian lengkap terkait nama *field*, tipe data, serta deskripsi fungsional setiap kolom disajikan dalam Tabel 3.8.

**Tabel 3. 8 Struktur Tabel *tbl\_dataset***

| Nama Field | Tipe Data | Panjang | Keterangan                 |
|------------|-----------|---------|----------------------------|
| <i>id</i>  | int       | 11      | ID unik dokumen lirik (PK) |

|                  |           |     |                                       |
|------------------|-----------|-----|---------------------------------------|
| id_admin         | int       | 11  | ID pengelola sistem (FK)              |
| artist           | varchar   | 255 | Nama artis pengubah lagu              |
| title            | varchar   | 255 | Judul lagu                            |
| genre            | varchar   | 100 | Kategori genre musik lagu tersebut    |
| lyrics           | longtext  |     | Data teks mentah dari lirik lagu      |
| case_folding     | text      |     | Lirik hasil konversi huruf kecil      |
| cleaning         | text      |     | Lirik bebas angka dan tanda baca      |
| tokenizing       | text      |     | Teks lirik yang dipecah menjadi token |
| stopword_removed | text      |     | Teks bebas noise dan kata hubung      |
| stemmed          | text      |     | Lirik dalam bentuk kata dasar         |
| upload_date      | timestamp |     | Waktu data diunggah ke sistem         |
| clean_lyrics     | text      |     | Lirik bersih untuk proses mesin       |
| sentiment_label  | varchar   | 20  | Target kelas prediksi aktual          |
| sentiment_result | varchar   | 20  | Hasil tebakan dari Naive Bayes        |

### 3.8 Rancangan Pengujian

Tahapan pengujian dirancang secara sistematis guna mengevaluasi efektivitas sistem klasifikasi sentimen lirik lagu, baik dari aspek fungsionalitas perangkat lunak maupun akurasi pemodelan bahasa alami terhadap data korpus lirik. Proses evaluasi ini dijalankan melalui implementasi aplikasi berbasis web yang mengintegrasikan basis data tugasakhir\_andika ke dalam satu ekosistem penyimpanan. Integrasi tersebut difungsikan untuk memfasilitasi eksekusi seluruh rangkaian proses, mulai dari *preprocessing*, ekstraksi fitur pembobotan TF-IDF, hingga penerapan algoritma *Multinomial Naive Bayes*.

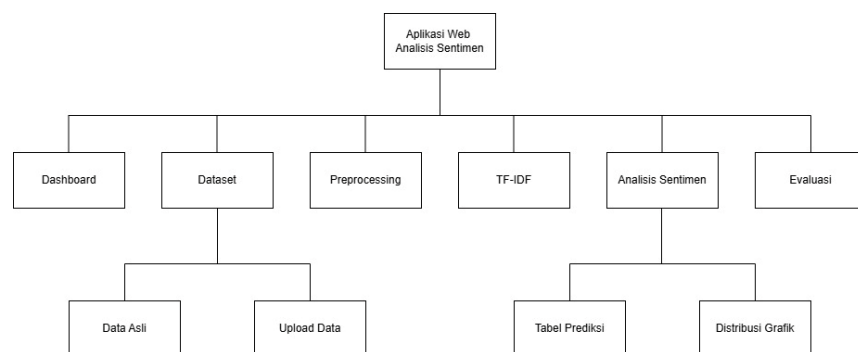
Evaluasi kinerja sistem dibagi menjadi dua dimensi utama. Pertama, penilaian fungsionalitas sistem (*Blackbox Testing*) dilakukan tanpa melibatkan struktur kode internal aplikasi, melainkan berfokus pada pengujian interaksi antarmuka pengguna seperti kelancaran proses unggah *dataset* CSV, eksekusi pembersihan teks, dan navigasi menu. Penggunaan metode pengujian ini bertujuan untuk mengukur tingkat kehandalan dan stabilitas sistem sebelum dioperasikan secara penuh.

Kedua, validasi performa model komputasi dilakukan dengan mengomparasikan hasil tebakan kelas sentimen (positif, netral, negatif) yang dikeluarkan oleh mesin terhadap label sentimen aktual (*ground truth*) dari dokumen lirik tersebut. Indikator keberhasilan pada tahap ini diukur secara kuantitatif melalui implementasi *Confusion Matrix* menggunakan parameter *Accuracy*, *Precision*, *Recall*, serta *F1-Score*. Melalui pendekatan pengujian yang komprehensif ini, diharapkan dapat diperoleh gambaran utuh mengenai ketepatan sistem komputasi dalam mengenali polaritas emosi pada teks lirik musik R&B dan *neo-soul*.

### 3.9 Rancangan Menu

Rancangan menu merupakan representasi visual dari arsitektur informasi dan struktur navigasi antarmuka pada sistem yang akan dibangun. Perancangan ini dipetakan secara konseptual dalam bentuk diagram pohon hierarki (*sitemap*) yang berfungsi untuk memvisualisasikan seluruh fungsionalitas aplikasi secara logis dan terstruktur. Melalui rancangan navigasi ini, alur interaksi pengguna dan tahapan perpindahan antar halaman dapat dipahami dengan jelas. Hal tersebut menjadi esensial untuk memberikan panduan penggunaan yang intuitif, mulai dari tahap inisiasi data hingga visualisasi hasil akhir pemrosesan penambangan teks (*text mining*).

Dalam pengembangannya, struktur menu pada sistem analisis lirik lagu ini dirancang secara sekuensial guna mengakomodasi alur komputasi eksperimen secara utuh. Arsitektur sistem ini mencakup enam modul utama, yaitu *Dashboard*, *Dataset*, *Pre-processing*, *TF-IDF*, *Analisis Sentimen*, dan *Evaluasi Model*. Khusus pada modul *Dataset*, struktur dikembangkan menjadi sistem hierarki bertingkat (*dropdown*) yang memuat dua sub-menu fungsional, yaitu *Upload Dataset* dan *Data Lirik*. Pembagian dan pengelompokan menu tersebut memastikan bahwa setiap proses transformasi data tekstual memiliki ruang eksekusi yang spesifik dan terorganisasi dengan baik. Bentuk visual dari rancangan arsitektur menu sistem tersebut dapat dilihat pada Gambar 3.3 berikut:



**Gambar 3.3 Rancangan Menu**

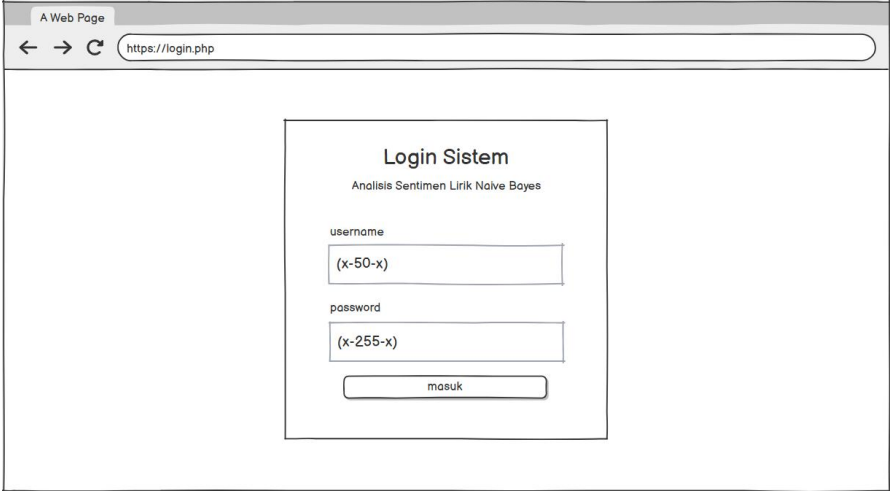
### 3.10 Rancangan Layar

Rancangan layar (*User Interface Design*) merupakan representasi visual dari antarmuka sistem web yang akan dibangun. Tujuan dari perancangan ini

adalah untuk memberikan gambaran arsitektur informasi dan tata letak (*layout*) elemen-elemen interaktif, sehingga pengoperasian fungsionalitas sistem oleh pengguna dapat berjalan secara intuitif dan efisien. Berikut adalah rancangan antarmuka untuk masing-masing halaman pada sistem klasifikasi sentimen lirik lagu *R&B* dan *Neo-Soul*:

### 3.10.1 Rancangan Layar Login

Halaman *login* merupakan gerbang autentikasi pertama bagi administrator untuk dapat mengakses sistem. Pada antarmuka ini, terdapat formulir yang mewajibkan pengguna untuk memasukkan kredensial berupa *username* dan *password* yang valid sebelum mendapatkan hak akses ke halaman utama.

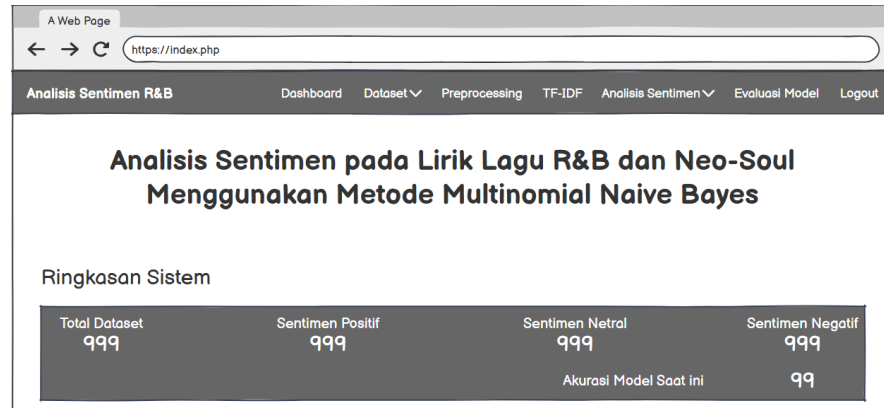


The image shows a web browser window with the address bar displaying 'https://login.php'. The main content area contains a login form titled 'Login Sistem' with the subtitle 'Analisis Sentimen Lirik Naive Bayes'. The form has two input fields: 'username' with a placeholder '(x-50-x)' and 'password' with a placeholder '(x-255-x)'. Below the fields is a button labeled 'masuk'.

Gambar 3. 4 Rancangan Layar Login

### 3.10.2 Rancangan Layar Dashboard

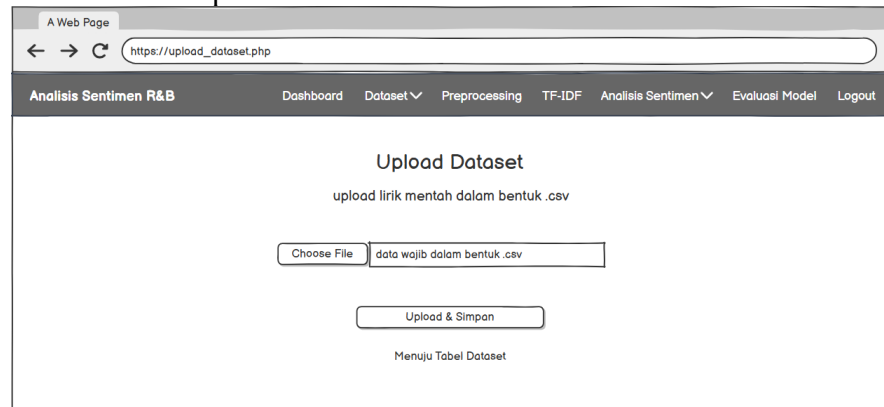
Halaman *dashboard* dirancang sebagai panel informasi utama yang pertama kali dilihat oleh admin setelah berhasil melewati proses autentikasi. Halaman ini berfungsi untuk menyajikan ringkasan singkat berupa indikator angka atau kartu informasi (*information cards*) terkait keseluruhan distribusi data yang ada di dalam basis data sistem.



**Gambar 3. 5 Rancangan Layar Dashboard**

### 3.10.3 Rancangan Layar Upload CSV

Halaman ini disediakan secara khusus sebagai modul masukan (*input*) data korpus. Admin dapat mencari, memilih, dan mengunggah berkas kumpulan lirik lagu dalam format CSV (*Comma Separated Values*) melalui tombol *Select File* sebelum data tersebut direkam secara permanen.



**Gambar 3. 6 Rancangan Layar Upload CSV**

### 3.10.4 Rancangan Layar Data Lirik Asli (*Raw Dataset*)

Halaman *Raw Dataset* berfungsi untuk menampilkan daftar dokumen lirik R&B dan Neo-Soul secara utuh yang belum mengalami proses modifikasi komputasional. Tabel pada halaman ini menyajikan detail informasi yang meliputi ID, judul lagu, nama artis, cuplikan lirik mentah (*snippet*), serta label sentimen manual yang menjadi acuan pengujian.

| ID    | Judul Lagu | Artis     | Lirik Mentah | Label Manual |
|-------|------------|-----------|--------------|--------------|
| q(11) | (x-255-x)  | (x-255-x) | (x-n-x)      | (x-20-x)     |
|       |            |           |              |              |
|       |            |           |              |              |
|       |            |           |              |              |
|       |            |           |              |              |

**Gambar 3. 7 Rancangan Layar Dataset Raw**

### 3.10.5 Rancangan Layar *Pre-Processing*

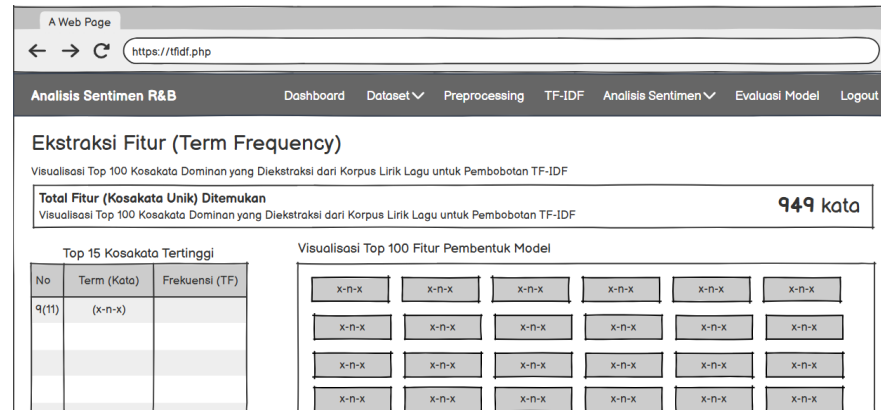
Halaman ini menyajikan tabel komparasi transparan mengenai rekam jejak pembersihan teks (*text preprocessing*). Melalui tabel ini, admin dapat memantau transformasi lirik lagu secara bertahap dari bentuk mentah hingga tereduksi melalui proses *case folding*, *cleaning*, *tokenizing*, dan *stopword removed*.

| No    | Info Lagu | Lirik Mentah | Case Folding | Cleaning | Tokenizing | Stopword Removal | Stemming |
|-------|-----------|--------------|--------------|----------|------------|------------------|----------|
| q(11) | (x-255-x) | (x-n-x)      | (x-n-x)      | (x-n-x)  | (x-n-x)    | (x-n-x)          | (x-n-x)  |
|       |           |              |              |          |            |                  |          |
|       |           |              |              |          |            |                  |          |
|       |           |              |              |          |            |                  |          |
|       |           |              |              |          |            |                  |          |

**Gambar 3. 8 Rancangan Layar Pre-processing**

### 3.10.6 Rancangan Layar Ekstraksi Fitur (TF-IDF)

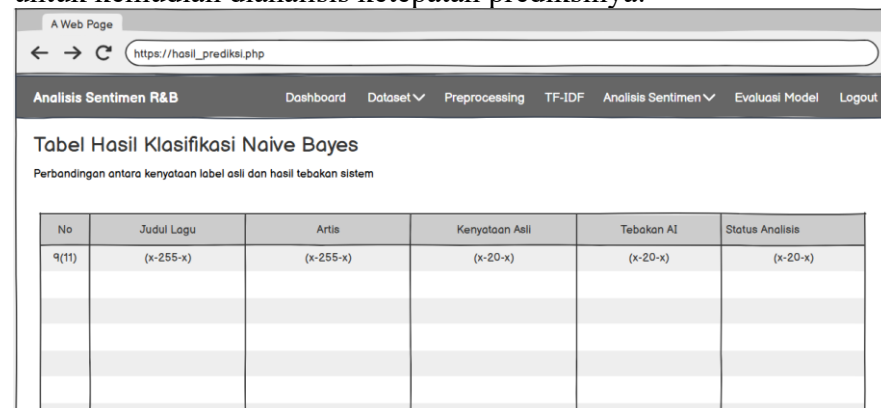
Halaman ekstraksi fitur memperlihatkan hasil kalkulasi pembobotan kata menggunakan algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF). Tampilan ini memuat informasi total kosakata unik yang berhasil diekstrak, tabel daftar kata dengan frekuensi kemunculan tertinggi, serta representasi visual dari fitur kata yang dominan dalam membentuk model.



Gambar 3. 9 Rancangan Layar Ekstraksi Fitur

### 3.10.7 Rancangan Layar Tabel Hasil Klasifikasi

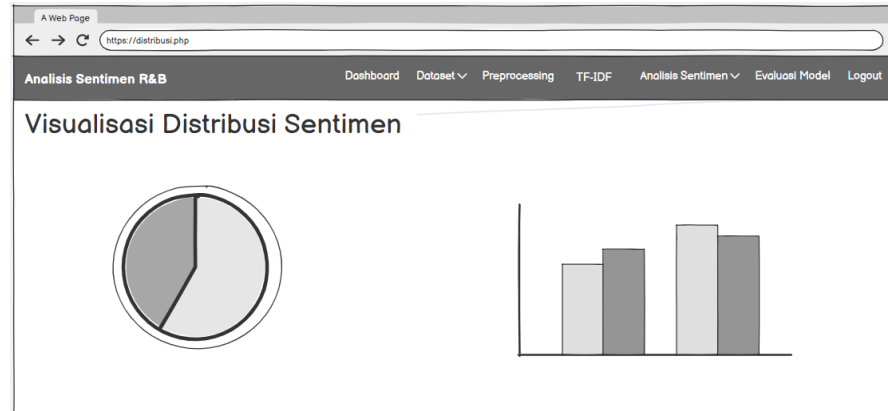
Halaman hasil klasifikasi mendeskripsikan luaran utama komputasi dari algoritma *Multinomial Naive Bayes*. Tabel yang disajikan memuat hasil prediksi polaritas sentimen terhadap dokumen teks lirik untuk kemudian dianalisis ketepatan prediksinya.



Gambar 3. 10 Rancangan Layar Tabel Hasil Klasifikasi

### 3.10.8 Rancangan Layar Visualisasi Distribusi Sentimen

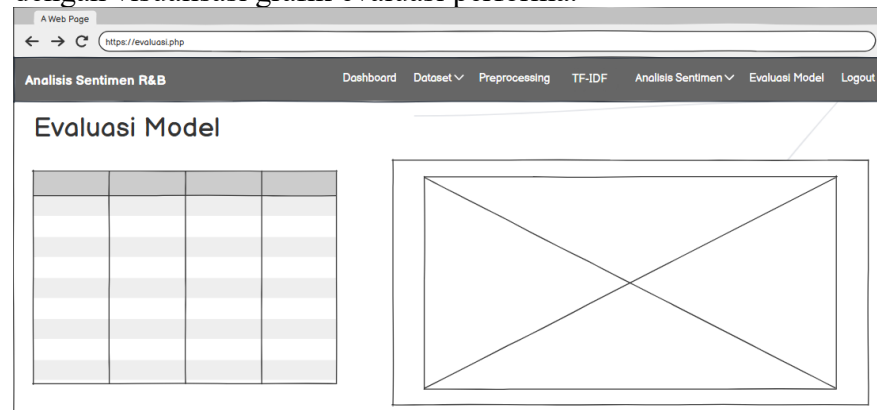
Sebagai pelengkap analisis prediktif, halaman ini merangkum sebaran hasil pelabelan sentimen ke dalam bentuk grafik visual. Terdapat representasi berupa diagram lingkaran (*pie chart*) dan diagram batang (*bar chart*) untuk memudahkan pembacaan komparasi antara kuantitas sentimen pada korpus.



**Gambar 3. 11 Rancangan Layar Visualisasi Distribusi Sentimen**

### 3.10.9 Rancangan Layar Evaluasi Model

Halaman terakhir ini merupakan panel evaluasi yang merangkum tingkat kehandalan algoritma klasifikasi. Antarmuka ini menampilkan parameter matriks kebingungan (*Confusion Matrix*) serta indikator metrik pengukuran kinerja model, yang didampingi dengan visualisasi grafik evaluasi performa.



**Gambar 3. 12 Rancangan Layar Evaluasi Model**

## BAB IV HASIL DAN PEMBAHASAN

### 4.1 Lingkungan Percobaan

Lingkungan percobaan merupakan penjabaran mengenai spesifikasi perangkat keras (*hardware*) dan perangkat lunak (*software*) yang digunakan sebagai infrastruktur komputasi selama proses pengembangan sistem. Lingkungan ini mencakup alat yang digunakan untuk penulisan kode program, pengelolaan basis data, eksekusi algoritma *Machine Learning*, hingga antarmuka berbasis web.

#### 4.1.1 Spesifikasi Hardware

Implementasi sistem analisis sentimen ini dilakukan menggunakan perangkat keras berupa satu unit laptop dengan spesifikasi sebagai berikut:

- |                 |                           |
|-----------------|---------------------------|
| a. Processor    | : Intel Core i5-13420H    |
| b. RAM          | : 16GB                    |
| c. Storage      | : 512GB                   |
| d. Graphic Card | : NVIDIA GeForce RTX 2050 |

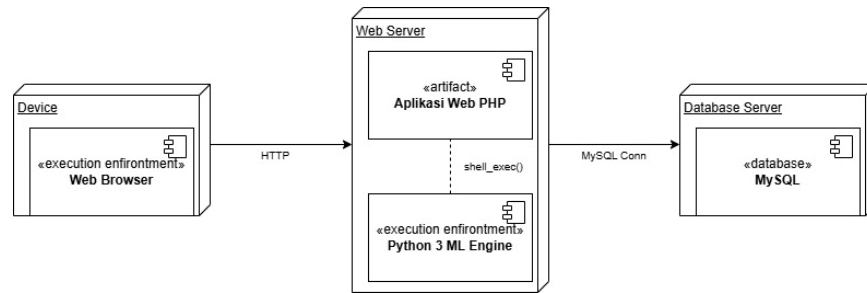
#### 4.1.2 Spesifikasi Software

Perangkat lunak yang digunakan dalam membangun dan menguji sistem ini adalah sebagai berikut:

- |                       |                                             |
|-----------------------|---------------------------------------------|
| a. IDE                | : Visual Studio Code                        |
| b. Browser            | : Microsoft Edge                            |
| c. Bahasa Pemrograman | : Python dan PHP                            |
| d. DBMS               | : MySQL                                     |
| e. OS                 | : Windows 11                                |
| f. Lainnya            | : XAMPP, Microsoft Excel, Draw.io, Balsamiq |

#### 4.1.3 Deployment Diagram

*Deployment diagram* berfungsi untuk memvisualisasikan arsitektur fisik penyebaran komponen perangkat lunak ke dalam simpul (*node*) perangkat keras. Pada sistem ini, arsitektur dibangun menggunakan model *client-server* yang beroperasi pada lingkungan lokal (*localhost*). *Browser* bertindak sebagai *client* yang mengirimkan permintaan (*request*), sementara Apache *Web Server* memproses logika PHP dan menjembatani eksekusi (*shell execution*) *script* Python untuk pemrosesan teks. Seluruh arsip data lirik dan hasil prediksi kemudian dikelola secara terpusat oleh server basis data MySQL.



**Gambar 4. 1 Deployment Diagram**

## 4.2 Implementasi Sistem

Implementasi sistem merupakan tahapan di mana rancangan konsep teoritis dan alur logika diwujudkan ke dalam bentuk operasional komputasi secara nyata. Pada tahapan ini, fokus pembahasan tidak ditekankan pada struktur kode pemrograman secara harfiah, melainkan pada penjabaran proses transformasi data (*Input-Process-Output*). Seluruh proses diurutkan mulai dari penyajian dataset mentah, pembersihan leksikal, ekstraksi matriks pembobotan, hingga hasil akhir prediksi algoritma dan penyimpanannya ke dalam basis data.

### 4.2.1 Implementasi Dataset

Implementasi dataset merupakan tahapan krusial pertama di mana korpus data lirik lagu berbasis teks diinisialisasi ke dalam ekosistem komputasi. Pada fase ini, sistem menerima masukan (input) berupa berkas berformat *Comma Separated Values* (.csv) yang memuat total keseluruhan 207 baris dokumen lirik bergenre R&B dan Neo-Soul.

Berkas dataset mentah tersebut memiliki struktur matriks dua dimensi yang telah dikurasi sebelumnya, mencakup beberapa atribut utama yang berfungsi sebagai parameter analisis awal. Adapun rincian atribut data tersebut adalah sebagai berikut:

- ID** : Merupakan atribut penanda identitas unik (nomor urut) untuk membedakan setiap dokumen lirik di dalam sistem.
- Artis** : Merupakan metadata yang memuat nama musisi, penyanyi, atau band pengubah lagu tersebut.
- Judul Lagu** : Merupakan metadata yang merepresentasikan identitas judul dari karya musik yang sedang dianalisis.
- Lirik Mentah** : Merupakan atribut inti yang memuat untaian teks murni dari keseluruhan bait lagu. Pada tahap ini, teks masih sepenuhnya bersifat tidak terstruktur (*unstructured data*), memiliki variasi kapitalisasi huruf, serta mengandung berbagai gangguan (*noise*) leksikal seperti tanda baca dan simbol.
- Label Manual** : Merupakan parameter target kelas polaritas aktual (Positif, Netral, atau Negatif) yang telah ditetapkan sebelumnya (*ground truth*). Atribut ini akan digunakan sebagai nilai acuan untuk memvalidasi ketepatan hasil prediksi algoritma komputasi pada tahap evaluasi model.

Setelah berkas dataset .csv diunggah melalui menu *Upload* pada antarmuka sistem web, mesin (*server*) akan membaca (*parsing*) keseluruhan struktur matriks tersebut dan menampilkannya sebagai basis data siap proses. Tabel 4.1 berikut mengilustrasikan sampel implementasi dari dataset lirik lagu mentah yang telah berhasil dimuat oleh sistem sebelum memasuki tahapan *pre-processing*:

**Tabel 4. 1 Sampel Implementasi Dataset Mentah**

| ID | Artis                | Judul Lagu         | Lirik Mentah (Raw Data)                                                      | Label Manual |
|----|----------------------|--------------------|------------------------------------------------------------------------------|--------------|
| 1  | Mac Ayres            | Easy               | It is easy to fall in love with you but hard to stay...                      | Positif      |
| 2  | xdika                | Teduh              | Langkahmu lirik namun, menyibak gelap tanpa ragu...                          | Positif      |
| 3  | Daniel Caesar        | Get You            | Through drought and famine natural disasters my baby...                      | Positif      |
| 4  | Glenn Fredly         | Sedih Tak Berujung | Saat menjelang hari hari bahagiamu, aku memilih diam dalam sepi... Oh, yeah! | Negatif      |
| 5  | Maliq & D'Essentials | Untitled           | Tidakkah cukup yang engkau lihat pertemanan ini sendiri...                   | Netral       |

#### 4.2.2 Implementasi Preprocessing

Setelah dataset mentah berhasil dimuat ke dalam sistem, tahapan selanjutnya adalah mengeksekusi proses *preprocessing* (pra-pemrosesan) pada atribut lirik lagu. Implementasi tahapan ini sangat krusial mengingat data lirik yang dikumpulkan masih berupa *unstructured data* yang mengandung banyak gangguan (*noise*) leksikal seperti variasi huruf kapital, tanda baca, simbol, imbuhan kata, serta kosakata umum yang tidak merepresentasikan polaritas sentimen.

Sistem mengimplementasikan lima tahapan *preprocessing* secara sekuensial menggunakan bahasa pemrograman Python, dengan memanfaatkan dukungan pustaka pemrosesan bahasa alami *Natural Language Toolkit* (NLTK) dan Sastrawi. Alur transformasi teks ini dieksekusi langkah demi langkah pada setiap baris dokumen lirik di dalam basis data untuk menghasilkan korpus kata yang bersih, padat, dan seragam.

Adapun lima tahapan yang diimplementasikan pada sistem meliputi:

##### a. *Case Folding*

Tahapan ini merupakan proses penyeragaman seluruh karakter huruf yang terdapat pada dokumen teks lirik lagu menjadi bentuk huruf

kecil (*lowercase*). Dalam pemrosesan bahasa alami, mesin komputasi bersifat sangat peka terhadap format huruf (*case-sensitive*), yang berarti huruf kapital dan huruf kecil dianggap sebagai nilai biner yang berbeda. Oleh karena itu, konversi ini wajib dilakukan untuk memastikan konsistensi pembacaan data, sehingga algoritma tidak mengalami kebingungan saat menghitung frekuensi kata yang sama namun memiliki gaya penulisan kapitalisasi yang berbeda di awal maupun di tengah kalimat.

*b. Cleaning*

Pada tahap pembersihan data ini, sistem secara sistematis menghapus seluruh karakter non-alfabetik dari dalam teks lirik yang tidak memiliki relevansi maupun nilai komputasi. Proses ini dieksekusi untuk mengeliminasi angka, karakter spesial, tautan, hingga berbagai bentuk tanda baca. Pembersihan karakter ini sangat esensial karena elemen-elemen tersebut sering kali hanya berfungsi sebagai pemanis tulisan atau *metadata* tambahan lagu (seperti tanda kurung pada lirik), namun tidak membawa muatan polaritas sentimen apapun bagi model algoritma dan justru berpotensi menjadi gangguan (*noise*).

*c. Tokenizing*

Setelah dokumen lirik dipastikan bersih dari karakter non-alfabetik, proses dilanjutkan dengan pemecahan untaian kalimat panjang menjadi potongan-potongan unit kata tunggal yang disebut sebagai *token*. Sistem mengeksekusi pemecahan ini dengan mengidentifikasi karakter spasi sebagai pemisah utamanya (*delimiter*). Tahapan ini bertindak sebagai jembatan yang sangat krusial sebelum memasuki fase pembobotan, karena algoritma TF-IDF dan *Naive Bayes* hanya dapat menganalisis dan mengkalkulasi probabilitas jika struktur teks tersebut telah diubah ke dalam bentuk susunan *array* kata yang terpisah.

*d. Stopword Removed*

Tahapan ini berfokus pada penyaringan tingkat lanjut untuk membuang kata-kata umum (*stopword*) yang memiliki tingkat frekuensi kemunculan sangat masif namun tidak menyumbangkan bobot informasi sentimen yang signifikan. Melalui implementasi pustaka Sastrawi dan NLTK, sistem mengeliminasi kata hubung, kata ganti, dan kata depan (seperti "aku", "di", "dan", "yang"). Sebagai bentuk optimasi khusus pada dataset lirik R&B dan Neo-Soul ini, sistem juga memberlakukan kamus *custom stopwords* untuk membuang *noise* musikal atau sisipan gumaman vokal (seperti "oh", "yeah", "mmm") guna menajamkan fitur kata dan mencegah terjadinya bias prediksi pada algoritma klasifikasi.

e. *Stemming*

Sebagai tahapan penutup pada fase *preprocessing*, *stemming* difungsikan untuk mentransformasikan token-token kata yang masih memiliki variasi imbuhan (afiksasi) kembali menjadi bentuk kata dasarnya (*root word*). Algoritma ini memotong prefiks maupun sufiks pada sebuah kata sehingga variasi leksikal diseragamkan menjadi satu bentuk dasar. Implementasi analisis morfologis ini sangat menguntungkan kinerja sistem komputasi secara keseluruhan karena sistem dapat melakukan reduksi dimensi matriks secara signifikan tanpa mengorbankan konteks makna sentimen dari lirik lagu yang dianalisis.

Untuk memberikan gambaran yang komprehensif mengenai rekam jejak transformasi bentuk teks pada tahapan ini, Tabel 4.2 berikut mengilustrasikan hasil implementasi *preprocessing* menggunakan sampel data lirik ID 4 (Glenn Fredly - Sedih Tak Berujung):

**Tabel 4. 2 Implementasi Tahapan *Preprocessing***

| Tahapan          | Output Hasil Transformasi Teks                                                                              | Keterangan                                                                       |
|------------------|-------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| Raw Text         | Saat menjelang hari hari bahagiamu, aku memilih diam dalam sepi... Oh, yeah!                                | Dokumen teks asli sebelum diproses.                                              |
| Case Folding     | saat menjelang hari hari bahagiamu, aku memilih diam dalam sepi... oh, yeah!                                | Seluruh huruf kapital dikonversi menjadi huruf kecil (lowercase).                |
| Cleaning         | saat menjelang hari hari bahagiamu aku memilih diam dalam sepi oh yeah                                      | Penghapusan seluruh karakter non alfabetik dan tanda baca.                       |
| Tokenizing       | ["saat", "menjelang", "hari", "hari", "bahagiamu", "aku", "memilih", "diam", "dalam", "sepi", "oh", "yeah"] | Pemecahan untaian kalimat teks lirik menjadi unit kata tunggal (token).          |
| Stopword Removal | ["menjelang", "hari", "hari", "bahagiamu", "memilih", "diam", "sepi"]                                       | Eliminasi kata hubung ("saat", "aku", "dalam") dan noise musikal ("oh", "yeah"). |
| Stemming         | ["jelang", "hari", "hari", "bahagia", "pilih", "diam", "sepi"]                                              | Transformasi token kata berimbuhan kembali ke bentuk kata dasar (root word).     |

Hasil akhir dari tahapan komputasi *preprocessing* ini akan menghasilkan atribut baru berupa lirik bersih (*clean lyrics*) di dalam tabel basis data. Teks lirik yang sudah bersih inilah yang nantinya akan dieksekusi pada tahapan pembobotan fitur dan prediksi *Machine Learning*.

#### 4.2.3 Implementasi TF-IDF

Tahapan implementasi TF-IDF (*Term Frequency - Inverse Document Frequency*) merupakan fase transformasi krusial dalam arsitektur sistem analisis sentimen ini. Algoritma *Machine Learning*, termasuk *Multinomial Naive Bayes*, pada dasarnya tidak memiliki kemampuan matematis untuk memproses dan mengkalkulasi data yang masih berwujud karakter alfabet murni. Oleh karena itu, sekumpulan teks lirik yang telah melewati fase pembersihan (*clean lyrics*) wajib dikonversi menjadi representasi vektor numerik berdimensi tinggi.

Implementasi ekstraksi fitur ini dieksekusi melalui modul pustaka komputasi ilmiah pada Python. Untuk membuktikan secara matematis proses ekstraksi fitur yang dilakukan oleh sistem, berikut adalah penjabaran perhitungan manual Term Frequency - Inverse Document Frequency (TF-IDF). Perhitungan ini mengambil contoh kata kunci "time" pada dokumen lirik lagu "*Japanese Denim*" karya Daniel Caesar dari total keseluruhan 207 dokumen lirik lagu yang ada di dalam dataset penelitian. Sistem memindai keseluruhan korpus data untuk membangun sebuah matriks kamus kata, di mana setiap *token* kata yang unik akan dievaluasi bobot urgensinya melalui dua tahapan perhitungan matematis utama:

a. *Term Frequency* (TF)

Sistem menghitung tingkat frekuensi kemunculan sebuah kata tertentu secara lokal di dalam satu baris dokumen lirik lagu. Semakin sering sebuah kata diulang dalam satu lagu, maka nilai *Term Frequency*-nya akan semakin tinggi. Frekuensi kemunculan kata "time" di dalam lirik lagu *Japanese Denim* (TF) = 4, kemudian total keseluruhan dokumen lirik dalam korpus (N) = 207, dan jumlah dokumen lagu yang mengandung kata "time" secara global (DF) = 80.

b. *Inverse Document Frequency* (IDF)

Sistem mengevaluasi tingkat keunikan kata secara global dengan menghitung jumlah dokumen lirik lain yang juga memuat kata tersebut. Fungsi ini bertindak sebagai penyeimbang dengan memberikan "penalti" atau penurunan bobot matematis pada kata-kata yang terlalu umum (muncul di hampir seluruh lagu dalam dataset). Sebaliknya, kata-kata spesifik yang menjadi ciri khas penentu sentimen akan mendapatkan dorongan bobot (*boosting*). Sistem menghitung nilai bobot penalti (IDF) untuk kata "time" menggunakan logaritma berbasis 10:

$$IDF = \log_{10} \left( \frac{N}{DF} \right)$$

$$IDF = \log_{10} \left( \frac{207}{80} \right)$$

$$IDF = \log_{10}(2.5875)$$

$$IDF \approx 0.413$$

Nilai bobot akhir (W) diperoleh dengan mengalikan nilai *Term Frequency* (TF) dengan nilai *Inverse Document Frequency* (IDF):

$$W_{time} = TF \times IDF$$

$$W_{time} = 4 \times 0.413 = 1.652$$

Hasil perkalian antara nilai TF dan IDF menghasilkan matriks bobot desimal akhir yang menunjukkan seberapa relevan dan berbobotnya sebuah kata kunci (*term*) terhadap narasi lagu tersebut. Nilai bobot desimal inilah yang nantinya akan disuplai sebagai data masukan (input) riil ke dalam model probabilitas *Naive Bayes*.

Untuk memberikan ilustrasi matematis mengenai hasil ekstraksi fitur tersebut, Tabel 4.3 menjabarkan simulasi hasil pembobotan TF-IDF pada beberapa *term* kata kunci yang diekstraksi dari korpus lirik R&B dan Neo-Soul di dalam sistem.

**Tabel 4. 3 Rekapitulasi Bobot TF-IDF pada Lagu Japanese Denim**

| <i>Term</i><br>(Kata Kunci) | <i>Term Frequency</i><br>(TF) | <i>Document Frequency</i><br>(DF) | <i>Inverse Document Frequency</i><br>(IDF) | Bobot TF IDF Akhir |
|-----------------------------|-------------------------------|-----------------------------------|--------------------------------------------|--------------------|
| time                        | 4                             | 80                                | 0.413                                      | 1.652              |
| love                        | 2                             | 100                               | 0.316                                      | 0.632              |
| say                         | 1                             | 70                                | 0.471                                      | 0.471              |
| jeans                       | 2                             | 2                                 | 2.015                                      | 4.030              |
| want                        | 0                             | 67                                | 0.490                                      | 0                  |

#### 4.2.4 Implementasi Multinomial Naive Bayes

Setelah seluruh dokumen teks diekstraksi dan dikonversi menjadi representasi matriks bobot numerik melalui algoritma TF-IDF, tahap komputasi inti selanjutnya adalah proses pengklasifikasian polaritas sentimen. Pada sistem ini, pengklasifikasian dieksekusi menggunakan model pembelajaran mesin terawasi (*supervised learning*) dengan mengimplementasikan algoritma probabilitas varian *Multinomial Naive Bayes* (MNB). Pemilihan varian multinomial dalam arsitektur sistem ini didasarkan pada kecocokannya secara matematis untuk menangani fitur

data diskrit berupa perhitungan frekuensi, sehingga sangat ideal untuk disandingkan dengan matriks keluaran dari pembobotan lirik lagu. Secara mekanis, implementasi klasifikasi algoritma ini berjalan melalui dua fase komputasi utama di dalam sistem, yaitu:

- a. Fase Pembelajaran (*Training*)  
Sistem memproses sekumpulan korpus data latih (*training data*) yang sebelumnya telah diberi label sentimen manual (*ground truth*). Pada tahap ini, algoritma membangun basis pengetahuan (*knowledge base*) dengan menghitung *prior probability* (peluang kemunculan keseluruhan kelas sentimen: Positif, Netral, dan Negatif) serta nilai *likelihood* (probabilitas kondisional dari keberadaan masing-masing fitur kata spesifik terhadap setiap kelas sentimen).
- b. Fase Klasifikasi (*Training/Prediksi*)  
Ketika sistem mengeksekusi proses klasifikasi pada data lirik uji, algoritma akan mengalkulasi nilai *posterior probability* (probabilitas akhir) dari dokumen lirik tersebut berdasarkan kemunculan matriks kata pembentuknya. Merujuk pada prinsip independensi Teorema Bayes, sistem secara otomatis akan melabeli dokumen lirik tersebut dengan kelas sentimen yang mengantongi nilai probabilitas agregat tertinggi.

Melalui pendekatan pemodelan statistik ini, lirik lagu genre R&B dan Neo-Soul yang awalnya bersifat ambigu dapat diekstraksi kecenderungan emosinya secara objektif oleh mesin. Untuk memberikan visualisasi luaran dari pemrosesan kecerdasan buatan tersebut, Tabel 4.4 menyajikan sampel implementasi komparasi antara hasil tebakan akhir mesin (Prediksi Sistem) dengan sentimen sesungguhnya (Label Aktual).

**Tabel 4. 4 Implementasi Prediksi Klasifikasi Sentimen**

| ID | Judul Lagu         | Lirik Bersih (Clean Lyrics)                    | Label Aktual | Prediksi Sistem | Status Klasifikasi |
|----|--------------------|------------------------------------------------|--------------|-----------------|--------------------|
| 1  | Easy               | easy fall love<br>hard stay                    | Positif      | Positif         | True               |
| 2  | Teduh              | langkah lirik<br>sibak gelap ragu              | Positif      | Positif         | True               |
| 3  | Get You            | drought famine<br>natural disaster<br>baby     | Positif      | Positif         | True               |
| 4  | Sedih Tak Berujung | jelang hari hari<br>bahagia pilih<br>diam sepi | Negatif      | Negatif         | True               |
| 5  | Untitled           | tidak cukup<br>engkau lihat<br>teman sendiri   | Netral       | Negatif         | False              |

(Catatan: Penentuan status True/False dilakukan dengan membandingkan kesesuaian antara parameter Label Aktual dan Prediksi Sistem. Jika kelas

sentimen yang diprediksi oleh algoritma tepat sasaran, maka klasifikasi dinyatakan True).

Untuk memberikan transparansi terhadap cara kerja algoritma *Multinomial Naive Bayes*, berikut dijabarkan simulasi perhitungan manual dalam memprediksi kelas sentimen pada lirik lagu "*Japanese Denim*". Proses klasifikasi menggunakan persamaan dasar Teorema Bayes dengan penambahan *Laplace Smoothing* (+1) untuk mencegah terjadinya probabilitas nol (0) pada kata yang tidak terdeteksi di dalam dokumen tersebut (seperti kata "*want*" dan "*feel*").

Sistem pertama-kali menghitung peluang kemunculan kelas secara global (*Prior Probability*). Dari total keseluruhan 207 dataset, diketahui distribusi sentimen terdiri dari 100 lagu Positif, 57 lagu Netral, dan 50 lagu Negatif. Berdasarkan data tersebut, peluang *Prior* untuk kelas Positif dihitung dengan perbandingan jumlah kelas terhadap total korpus:

$$P(\text{Positif}) = \frac{100}{207} \approx 0.483$$

Tahap selanjutnya adalah menghitung probabilitas kondisional (*Likelihood*) dari setiap kata kunci yang telah diekstraksi bobot TF-IDF-nya (seperti "*time*", "*love*", "*say*", "*jeans*") terhadap masing-masing kelas. Sebagai contoh perhitungan pada kelas Positif, sistem memindai bahwa total keseluruhan kata yang menyusun kelas Positif ( $N_c$ ) adalah sebanyak 36.731 kata, dengan total kosakata unik secara global ( $|V|$ ) sebanyak 4.057 kata. Berdasarkan pemindaian matriks, frekuensi kemunculan kata "*time*" di dalam seluruh lagu bersentimen Positif ( $N_{w,c}$ ) tercatat sebanyak 148 kali. Dengan menggunakan penambahan *Laplace Smoothing*, nilai *Likelihood* kata "*time*" untuk kelas Positif dihitung sebagai berikut:

$$P(\text{"time"}|\text{Positif}) = \frac{148 + 1}{36731 + 4057} = \frac{149}{40788} \approx 0.00365$$

Sistem komputasi kemudian mengulangi siklus perhitungan probabilitas kondisional yang sama untuk seluruh matriks kata pembentuk lagu "*Japanese Denim*" pada kelas Netral dan Negatif. Pada tahap penentuan keputusan, seluruh nilai *Likelihood* dari matriks kata tersebut diagregasikan dan dikalikan dengan nilai *Prior Probability* untuk menghasilkan probabilitas akhir (*Posterior Probability*). Kelas sentimen yang mengantongi nilai *Posterior* agregat tertinggi akan dieksekusi secara otomatis sebagai hasil prediksi akhir (Tebakan AI) dari sistem.

**Tabel 4. 5 Rekapitulasi Probabilitas Akhir Naive Bayes pada Lagu *Japanese Denim***

| Kelas Sentimen | <i>Prior Probability</i><br>$P(C)$ | Agregat<br><i>Likelihood</i><br>$P(X  C)$ | Posterior<br>Probability | Hasil Prediksi Sistem |
|----------------|------------------------------------|-------------------------------------------|--------------------------|-----------------------|
| Positif        | 0.483                              | 0.00125                                   | 0.00060                  | Terpilih (Tertinggi)  |
| Netral         | 0.275                              | 0.00042                                   | 0.00011                  | -                     |
| Negatif        | 0.242                              | 0.00018                                   | 0.00004                  | -                     |

#### 4.2.5 Implementasi Basis Data

Implementasi basis data merupakan tahap akhir dari siklus arsitektur *input-process-output* pada sistem. Seluruh rangkaian eksekusi teks, mulai dari penyimpanan data awal, histori *preprocessing*, hingga rekam jejak klasifikasi sentimen diintegrasikan secara persisten ke dalam sistem manajemen basis data relasional (RDBMS) MySQL.

Pada arsitektur sistem ini, seluruh aliran data dikelola secara terpusat di dalam basis data bernama tugasakhir\_andika, yang dikonstruksi atas dua entitas tabel utama, yaitu:

a. Entitas (tbl\_admin)

Tabel ini difungsikan secara tertutup (*backend*) untuk menampung data kredensial autentikasi. Data *username* dan *password* dari administrator sistem direkam pada entitas ini guna memberikan pembatasan hak akses yang aman untuk pengelolaan dataset lirik lagu.

b. Entitas (tbl\_dataset)

Tabel ini bertindak sebagai repositori utama dan basis pengetahuan (*knowledge base*) komputasi sistem. Setiap kali sistem mengeksekusi tahapan analisis, *Machine Learning Engine* akan merekam luaran dari masing-masing proses komputasi Python dan menyimpannya ke dalam atribut kolom yang spesifik. Lirik lagu mentah disimpan pada atribut *lyrics*, sementara histori *preprocessing* direkam secara bertahap pada atribut *case\_folding*, *cleaning*, *tokenizing*, *stopword\_removed*, hingga lirik bersih pada atribut *stemmed* dan *clean\_lyrics*. Lebih lanjut, label sentimen manual beserta hasil prediksi akhir dari *Naive Bayes* direkam secara permanen pada atribut *sentiment\_label* dan *sentiment\_result*.

Implementasi arsitektur penyimpanan yang terstruktur ini menjamin bahwa seluruh proses penambangan teks (*text mining*) tidak berjalan secara temporal (hilang saat sesi *browser* berakhir), melainkan tersimpan secara permanen. Data yang telah direkam secara komprehensif pada *tbl\_dataset* inilah yang kemudian dipanggil (*di-query*) menggunakan bahasa

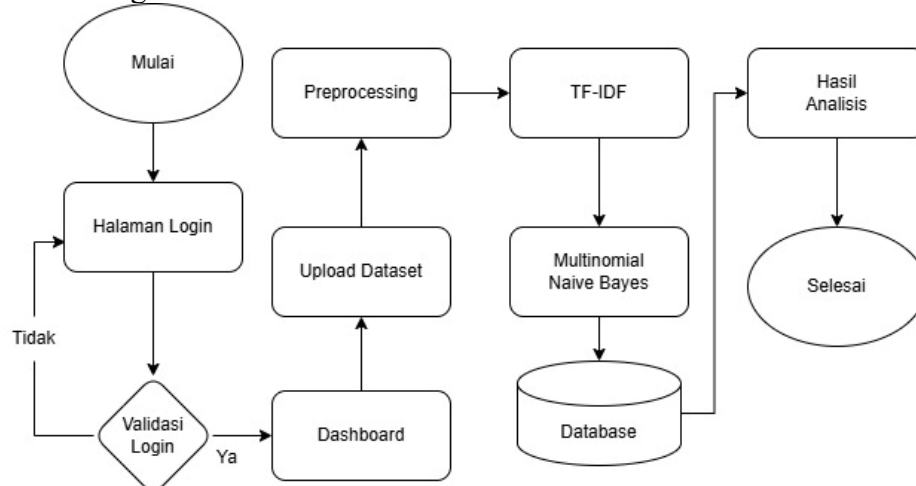
pemrograman PHP untuk divisualisasikan secara *real-time* ke dalam bentuk antarmuka tabel laporan dan grafik distribusi pada *dashboard* aplikasi web.

#### 4.3 Flowchart Sistem

*Flowchart* atau diagram alir merupakan representasi visual yang memetakan urutan langkah-langkah logis dan keputusan operasional dalam sebuah sistem komputasi. Pada penelitian ini, *flowchart* digunakan untuk menggambarkan alur kerja arsitektur sistem analisis sentimen lirik lagu secara terstruktur, guna memudahkan pemahaman mengenai bagaimana data lirik diproses dari wujud mentah hingga menjadi luaran prediksi yang informatif.

##### 4.3.1 Flowchart Keseluruhan Sistem

*Flowchart* keseluruhan sistem memvisualisasikan alur interaksi makro antara pengguna (administrator) dengan sistem aplikasi web. Proses diawali dengan tahapan autentikasi keamanan, dilanjutkan dengan inisialisasi korpus teks, rangkaian komputasi penambangan teks (*text mining*), hingga penyajian luaran akhir berupa tabel evaluasi dan visualisasi grafik sentimen.



**Gambar 4. 2 Flowchart Keseluruhan Sistem**

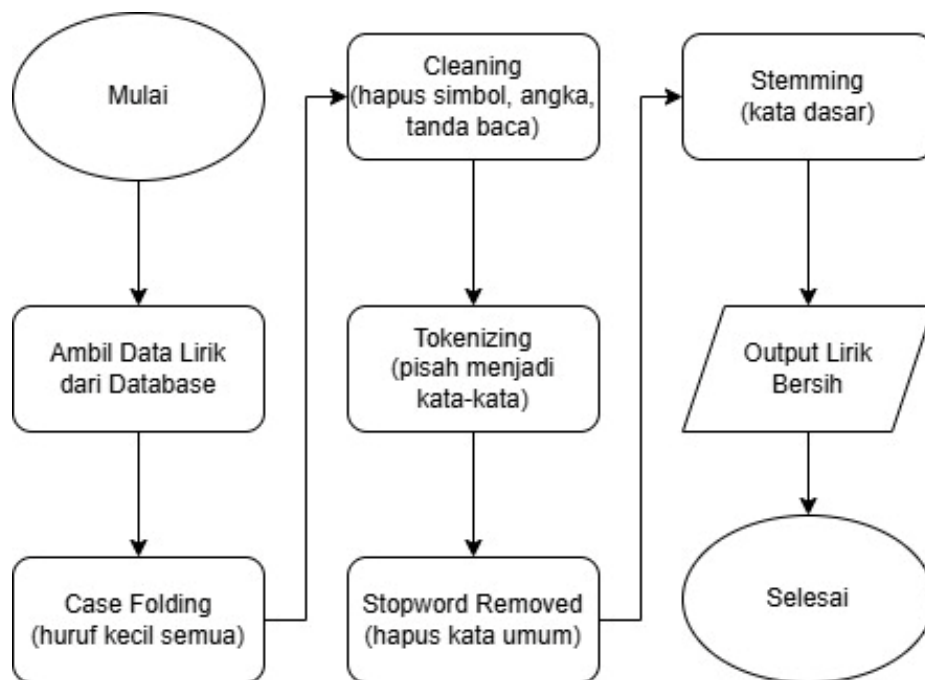
Berdasarkan diagram alir keseluruhan sistem pada Gambar 4.2 di atas, alur kerja aplikasi diawali dengan tahapan inisiasi program yang menampilkan antarmuka halaman *login* kepada administrator. Pada tahap ini, administrator diwajibkan untuk menginputkan kredensial berupa *username* dan *password*. Sistem kemudian akan melakukan proses validasi data. Apabila kredensial yang dimasukkan salah, sistem akan menampilkan pesan peringatan dan mengembalikan administrator ke halaman *login*. Namun, jika validasi berhasil, sistem akan memberikan hak akses dan mengarahkan administrator langsung ke halaman *Dashboard*.

Setelah berhasil masuk, administrator dapat mengunggah berkas dataset lirik lagu mentah dalam format *.csv*. Data yang diunggah tersebut kemudian dibaca oleh sistem untuk dieksekusi ke dalam tahapan

*preprocessing* guna membersihkan teks lirik dari berbagai *noise*. Lirik yang telah bersih selanjutnya diproses ke dalam modul ekstraksi fitur untuk dilakukan perhitungan pembobotan TF-IDF. Matriks bobot angka yang dihasilkan kemudian diumpankan ke dalam algoritma *Multinomial Naive Bayes* untuk mengklasifikasikan polaritas sentimen ke dalam kelas Positif, Netral, atau Negatif. Setelah seluruh proses komputasi selesai, sistem akan merekam seluruh riwayat transformasi lirik beserta hasil prediksinya ke dalam *database* MySQL. Sebagai luaran akhir, sistem menyajikan hasil komputasi tersebut kepada administrator dalam bentuk tabel hasil klasifikasi, sebaran kata dominan, serta visualisasi grafik distribusi sentimen sebelum akhirnya proses dinyatakan selesai.

#### 4.3.2 Flowchart Preprocessing

*Flowchart preprocessing* memvisualisasikan alur pemrosesan secara spesifik pada tahap pembersihan dan transformasi teks lirik lagu. Karena data teks mentah (*raw data*) memiliki tingkat kompleksitas yang tinggi dan sarat akan *noise*, tahapan ini berjalan secara sekuensial (berurutan) agar *output* dari satu proses menjadi *input* bagi proses selanjutnya, hingga menghasilkan korpus lirik yang bersih dan terstandarisasi.



Gambar 4. 3 *Flowchart Preprocessing*

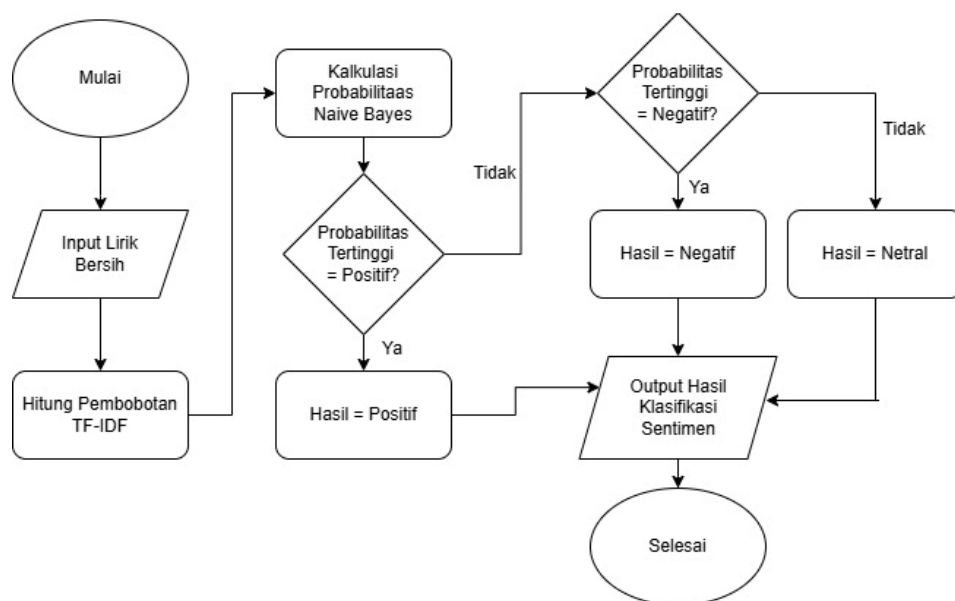
Berdasarkan diagram alir *preprocessing* pada Gambar 4.3 di atas, tahapan transformasi leksikal dimulai ketika sistem menerima masukan data berupa teks lirik mentah (*raw lyrics*). Data teks tersebut kemudian dieksekusi melalui serangkaian proses pembersihan yang berjalan secara sekuensial. Proses pertama adalah *case folding*, di mana sistem

menyeragamkan seluruh karakter teks menjadi huruf kecil. Setelah itu, sistem menjalankan tahap *cleaning* untuk mengeliminasi seluruh angka, simbol, serta tanda baca yang tidak relevan.

Teks yang telah dibersihkan kemudian dipecah menjadi matriks unit kata tunggal melalui tahapan *tokenizing*. Rangkaian *token* kata tersebut selanjutnya disaring dalam tahap *stopword removal* untuk membuang kata hubung dan *noise* musikal yang tidak memiliki bobot analitis. Tahap transformasi terakhir adalah *stemming*, di mana sistem mengonversi *token* kata yang masih memiliki variasi imbuhan kembali menjadi bentuk kata dasar. Seluruh rangkaian proses pembersihan ini pada akhirnya akan menghasilkan luaran berupa lirik bersih (*clean lyrics*) yang siap untuk diekstraksi bobotnya pada tahapan selanjutnya, yang sekaligus menandai berakhirnya proses komputasi *preprocessing*.

#### 4.3.3 Flowchart Klasifikasi Sentimen

*Flowchart* klasifikasi memvisualisasikan algoritma inti dari kecerdasan buatan pada sistem ini. Alur ini menggambarkan bagaimana dokumen lirik yang telah bersih (*clean lyrics*) diekstraksi nilainya dan diprediksi kelas sentimennya menggunakan probabilitas *Multinomial Naive Bayes*.



Gambar 4. 4 *Flowchart* Klasifikasi Sentimen

Berdasarkan diagram alir klasifikasi pada Gambar 4.4 di atas, tahapan prediksi sentimen diinisiasi saat sistem menerima input berupa dokumen lirik bersih (*clean lyrics*) yang telah melewati fase *preprocessing*. Sistem kemudian mengeksekusi ekstraksi fitur menggunakan pembobotan *Term Frequency - Inverse Document Frequency* (TF-IDF) untuk mentransformasikan teks tersebut menjadi representasi matriks bobot

angka. Matriks berbobot ini selanjutnya diumpankan ke dalam model algoritma *Multinomial Naive Bayes*. Pada tahap ini, sistem mengalkulasi probabilitas setiap fitur kata terhadap masing-masing kelas sentimen dengan merujuk pada data latih yang ada.

Setelah perhitungan probabilitas agregat akhir diperoleh, sistem akan melakukan evaluasi kondisi keputusan untuk menentukan label akhir lirik. Apabila nilai probabilitas tertinggi berada pada kelas Positif, maka dokumen lirik tersebut ditetapkan berlabel Positif. Sebaliknya, jika probabilitas tertinggi berada pada kelas Negatif, maka lirik ditetapkan sebagai Negatif. Namun, jika probabilitas tertinggi tidak memenuhi kedua kondisi tersebut—misalnya memiliki nilai yang seimbang—maka lirik akan diklasifikasikan ke dalam kelas Netral. Keputusan pembagian kelas sentimen inilah yang menjadi luaran akhir (*output*) dari proses komputasi klasifikasi sebelum alur algoritma dinyatakan selesai.

#### 4.4 Algoritma Sistem

Algoritma sistem merupakan penjabaran tekstual dari diagram alir (*flowchart*) yang memuat urutan langkah-langkah logis penyelesaian masalah secara komputasional. Pada tahap ini, alur kerja aplikasi diterjemahkan ke dalam bentuk *pseudocode* (kode semu) berstruktur. Pendekatan ini digunakan untuk memberikan pemahaman yang lebih teknis mengenai instruksi dasar yang akan dieksekusi oleh mesin, sebelum diimplementasikan secara harfiah ke dalam bahasa pemrograman Python dan PHP.

##### 4.4.1 Algoritma Keseluruhan Sistem

Algoritma keseluruhan sistem mendeskripsikan instruksi makro dari proses interaksi administrator dengan aplikasi berbasis web. Instruksi ini memuat logika autentikasi, inisiasi pengolahan data lirik, hingga penyajian luaran akhir. Berikut adalah penjabaran *pseudocode* dari algoritma keseluruhan sistem:

**Algoritma 4. 1 Proses Keseluruhan Sistem**

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Start<br>Tampilkan antarmuka halaman login<br>Baca input username dan password dari administrator<br>Jika kredensial salah, hentikan akses dan kembalikan ke halaman login<br>Jika kredensial benar, berikan hak akses dan tampilkan halaman dashboard<br>Unggah dan simpan data lirik lagu<br>Lakukan preprocessing pada data teks lirik<br>Lakukan ekstraksi fitur dan pembobotan menggunakan metode TF-IDF<br>Lakukan klasifikasi sentimen menggunakan algoritma Multinomial Naive Bayes<br>Simpan riwayat transformasi teks dan hasil klasifikasi ke dalam basis data |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                            |
|--------------------------------------------------------------------------------------------|
| Kembalikan luaran berupa tabel hasil klasifikasi dan grafik distribusi sentimen<br>Selesai |
|--------------------------------------------------------------------------------------------|

#### 4.4.2 Algoritma Preprocessing

Algoritma *preprocessing* menjabarkan urutan instruksi spesifik yang dieksekusi oleh sistem untuk membersihkan data teks lirik lagu dari berbagai gangguan (*noise*) leksikal. Rangkaian instruksi ini berjalan secara sekuensial untuk memastikan algoritma *Machine Learning* nantinya hanya memproses korpus kata yang bersih dan memiliki bobot informasi sentimen.

##### Algoritma 4. 2 Proses *Preprocessing*

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Start<br>Ambil data teks lirik mentah dari basis data<br>Lakukan case folding untuk mengubah seluruh karakter huruf menjadi huruf kecil<br>Lakukan cleaning untuk menghapus karakter angka, simbol, dan tanda baca<br>Lakukan tokenizing untuk memisahkan untaian kalimat menjadi potongan unit kata tunggal<br>Lakukan stopwords removal untuk membuang kata hubung dan noise musikal yang tidak bermakna<br>Lakukan stemming untuk mentransformasikan token kata berimbuhan kembali menjadi bentuk kata dasar<br>Hasilkan keluaran berupa teks lirik bersih<br>Selesai |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

#### 4.4.3 Algoritma Klasifikasi Sentimen

Algoritma klasifikasi sentimen mendeskripsikan urutan komputasi cerdas dari sistem dalam menentukan polaritas emosi pada lirik lagu. Instruksi ini memuat logika ekstraksi bobot kata menggunakan TF-IDF yang kemudian diumpankan ke dalam model probabilitas *Multinomial Naive Bayes* untuk mengambil keputusan akhir berupa penetapan kelas sentimen.

##### Algoritma 4. 3 Proses Klasifikasi Sentimen

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Start<br>Ambil data lirik bersih hasil preprocessing<br>Lakukan ekstraksi fitur dan pembobotan matematis menggunakan TF-IDF<br>Kalkulasi probabilitas setiap fitur kata terhadap data latih menggunakan Multinomial Naive Bayes<br>Tentukan nilai probabilitas agregat akhir yang paling tinggi<br>Jika probabilitas tertinggi mengarah pada kelas Positif, maka tetapkan dokumen berlabel Positif<br>Jika probabilitas tertinggi mengarah pada kelas Negatif, maka tetapkan dokumen berlabel Negatif |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Jika probabilitas tertinggi tidak memenuhi keduanya, maka tetapkan dokumen berlabel Netral  
Hasilkan keluaran berupa status kelas sentimen lirik akhir  
Selesai

#### 4.5 Pengujian dan Evaluasi

Tahapan pengujian dan evaluasi merupakan fase krusial untuk mengukur tingkat kehandalan dan validitas dari arsitektur sistem yang telah dibangun. Pada tahap ini, kemampuan algoritma *Multinomial Naive Bayes* diukur secara kuantitatif dalam memprediksi kelas polaritas sentimen (Positif, Netral, dan Negatif) dari korpus lirik lagu R&B dan Neo-Soul.

##### 4.5.1 Skenario Pengujian

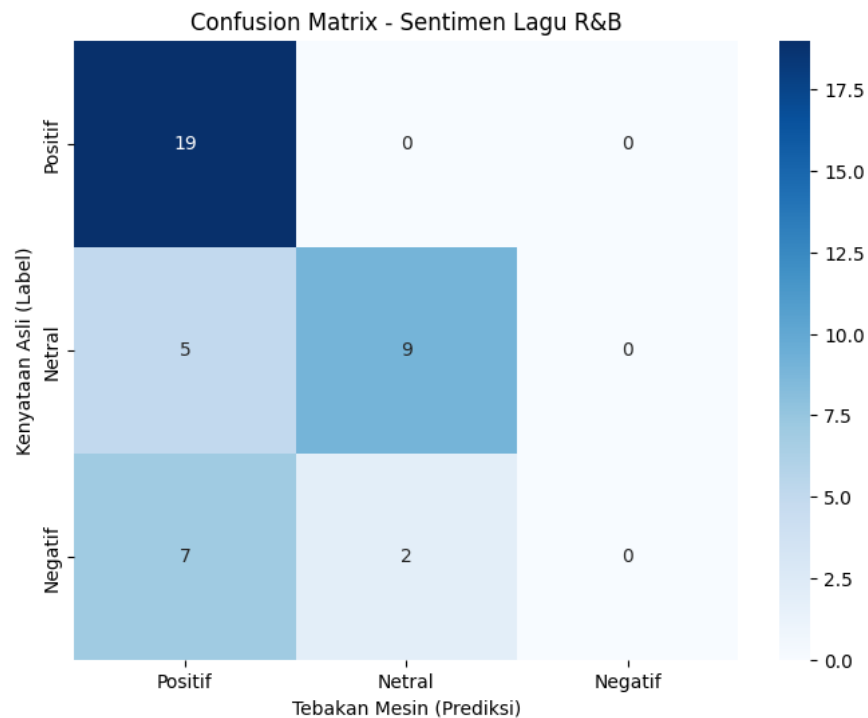
Skenario pengujian dirancang menggunakan metode *Data Splitting* (pembagian data) untuk memisahkan fungsi dataset ke dalam dua lingkungan komputasi yang berbeda: data latih (*training data*) dan data uji (*testing data*). Skenario ini diimplementasikan untuk memastikan bahwa mesin dievaluasi menggunakan dokumen lirik yang belum pernah ia "lihat" sebelumnya, sehingga hasil akurasi yang diperoleh merepresentasikan kemampuan prediksi yang sesungguhnya.

Pada penelitian ini, total korpus lirik sebanyak 207 dokumen dibagi menggunakan rasio standar **80:20**. Rincian distribusi datanya adalah sebagai berikut:

- a. Data Latih (*Training Data*) sebesar 80% (165 dokumen lirik):  
Digunakan untuk membangun basis pengetahuan (*knowledge base*) algoritma *Multinomial Naive Bayes*. Pada tahap ini, mesin mengalkulasi *prior probability* dan *likelihood* dari masing-masing fitur kata terhadap kelas sentimennya (Positif: 100, Netral: 57, Negatif: 50).
- b. Data Uji (*Testing Data*) sebesar 20% (42 dokumen lirik):  
Digunakan murni sebagai soal evaluasi (parameter *blind test*). Sistem akan menebak kelas sentimen dari 42 lirik ini, kemudian hasil tebakannya (*sentiment result*) akan dikomparasikan dengan kenyataan aslinya (*sentiment\_label / ground truth*).

##### 4.5.2 Evaluasi Model Klasifikasi

Pengukuran kinerja algoritma komputasi dievaluasi menggunakan pendekatan *Confusion Matrix* (Matriks Kebingungan). Matriks ini memetakan perbandingan langsung antara hasil tebak mesin dengan label sentimen aktual pada 42 data uji. Visualisasi hasil *Confusion Matrix* dari sistem disajikan pada Gambar 4.5.



**Gambar 4. 5 Visualisasi Hasil *Confusion Matrix***

Berdasarkan matriks pada Gambar 4.5 di atas, sebaran tebakan dari 42 data uji dapat dijabarkan sebagai berikut:

- Dari 19 data yang aslinya Positif, mesin berhasil menebak benar 19 data sebagai Positif (100% *True Positive*).
- Dari 14 data yang aslinya Netral, mesin menebak 9 data dengan benar sebagai Netral, namun 5 data lainnya meleset ditebak sebagai Positif.
- Dari 9 data yang aslinya Negatif, mesin gagal menebak seluruhnya (0). Mesin justru menebak 7 data sebagai Positif dan 2 data sebagai Netral.

Merujuk pada sebaran prediksi tersebut, sistem kemudian mengkalkulasi metrik performa algoritma secara keseluruhan. Nilai Akurasi (*Accuracy*) akhir yang berhasil dicapai oleh model *Multinomial Naive Bayes* dalam penelitian ini adalah sebesar **57.14%**. Nilai ini merepresentasikan bahwa sistem mampu memprediksi lebih dari separuh dokumen lirik dengan tepat.

Jika dibedah berdasarkan detail metrik per kelas:

- Kelas **Positif** meraih nilai tertinggi dengan *Precision* 0.58, *Recall* 0.95, dan *F1-Score* 0.72. Tingginya *Recall* membuktikan sistem sangat sensitif dalam mengenali lirik bernuansa positif.
- Kelas **Netral** memperoleh *Precision* 0.71, *Recall* 0.42, dan *F1-Score* 0.53.
- Kelas **Negatif** meraih nilai 0.00 pada semua metrik. Hal ini mengindikasikan adanya kendala pada model (*overfitting* pada

kelas mayoritas) dalam membedakan lirik sedih/negatif akibat struktur bahasa R&B yang sangat ambigu dan sarat akan *noise* leksikal.

#### 4.5.3 Analisis Hasil Sentimen

Selain melakukan pengujian akurasi pada sebagian data, sistem juga melakukan prediksi sentimen secara menyeluruh terhadap total 207 dokumen lirik R&B dan Neo-Soul di dalam basis data untuk melihat kecenderungan (tren) secara agregat. Berdasarkan pemrosesan melalui halaman *dashboard* aplikasi, diperoleh persebaran distribusi prediksi mesin sebagai berikut:

- a. Sentimen Positif: 168 lagu (Mayoritas dominan)
- b. Sentimen Netral: 31 lagu
- c. Sentimen Negatif: 8 lagu

Hasil komputasi ini menunjukkan adanya ketimpangan prediksi (*skewness*) di mana algoritma *Multinomial Naive Bayes* sangat agresif dalam melabeli dokumen ke dalam kelas Positif. Secara musikal dan linguistik, anomali ini dapat dibenarkan. Genre R&B dan Neo-Soul memiliki karakteristik penulisan lirik yang sangat emosional. Lagu yang secara manual berlabel Negatif (seperti lagu tentang perpisahan atau rasa kehilangan) sering kali tetap menggunakan kosakata afeksi tingkat tinggi. Kesamaan irisan leksikal antara lagu sedih dan lagu bahagia inilah yang menyebabkan probabilitas matematis mesin cenderung membelokkan hasil prediksi akhir menuju kelas Positif.

#### 4.5.4 Analisis Kata Kunci Dominan

Untuk memvalidasi alasan di balik ketimpangan prediksi tersebut, sistem mengeksekusi ekstraksi frekuensi term guna melihat kata kunci (*keywords*) apa saja yang paling dominan muncul pada masing-masing kelas sentimen. Analisis kata kunci ini menjadi representasi matematis dari cara mesin memahami makna bahasa kiasan dalam musik.

Berdasarkan ekstraksi fitur pembobotan pada sistem, berikut adalah sebaran kata kunci dominan:

- a. Kata Kunci Sentimen Positif  
Kumpulan lirik bernuansa positif sangat didominasi oleh kata-kata afirmasi dan afeksi universal, dengan frekuensi kemunculan tertinggi pada token kata seperti "*love*", "*baby*", "*feel*", "*want*", dan "*give*". Dalam lirik berbahasa Indonesia, muatan positif ini sering dikemas dalam diksi yang merepresentasikan kehadiran yang menenangkan hati, seperti "*teduh*", "*tenang*", dan "*bahagia*". Algoritma sangat sensitif dalam menangkap token-token ini sebagai bobot *True Positive*.
- b. Kata Kunci Sentimen Negatif  
Pada kelas polaritas negatif, sistem mendeteksi tingginya frekuensi pada token kata yang merepresentasikan penolakan, keterpurukan, maupun pertanyaan, seperti "*cant*", "*aint*",

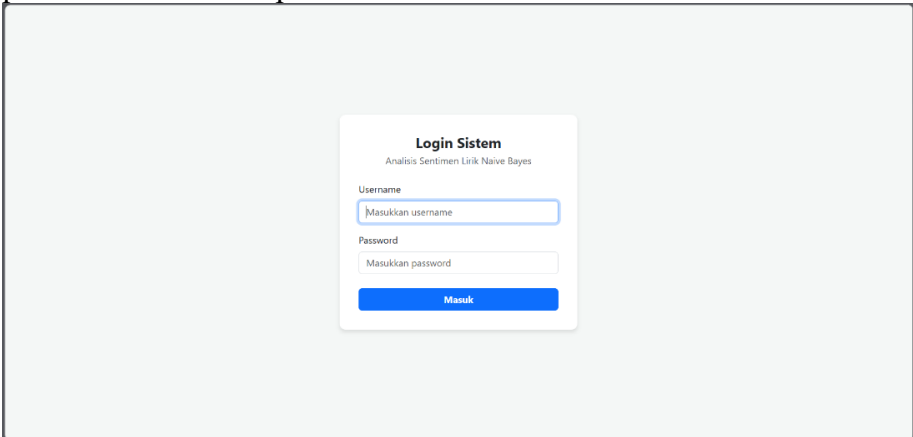
"down", dan "why". Namun, yang menjadi temuan menarik dalam analisis ini adalah kata "love" tetap muncul dengan frekuensi yang masif pada lagu bersentimen negatif. Keberadaan token "love" pada kalimat patah hati (contoh: "I can't love you anymore") inilah yang menjadi *noise* leksikal terbesar, mengelabui mesin untuk mengklasifikasikan dokumen tersebut sebagai sentimen Positif.

## 4.6 Tampilan Layar Aplikasi

Bagian ini menyajikan visualisasi antarmuka pengguna (*User Interface*) dari sistem aplikasi web analisis sentimen yang telah dibangun. Antarmuka ini dirancang secara interaktif untuk memudahkan administrator dalam melakukan pengelolaan korpus lirik lagu hingga memantau hasil klasifikasi algoritma *Multinomial Naive Bayes*.

### 4.6.1 Login

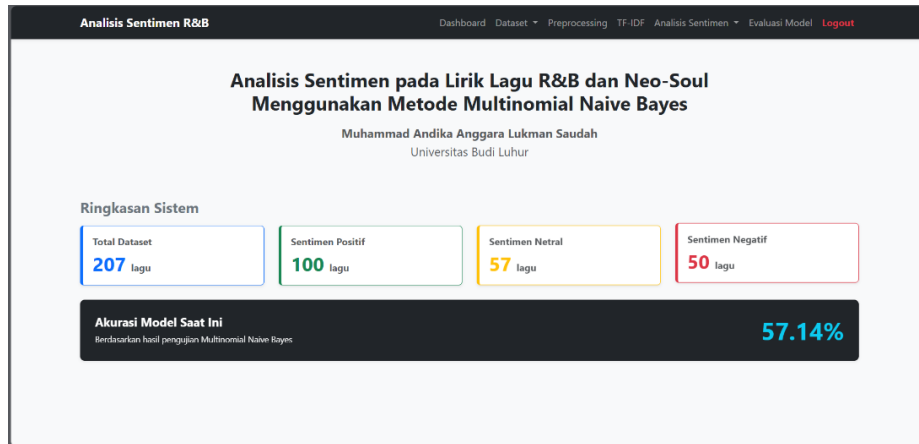
Halaman *login* merupakan gerbang autentikasi utama sistem. Pada halaman ini, administrator diwajibkan untuk memasukkan kredensial berupa *username* dan *password* yang valid guna mendapatkan hak akses penuh ke dalam fitur pemrosesan data.



Gambar 4. 6 Tampilan *Layar Login*

### 4.6.2 Dashboard

Halaman *dashboard* merupakan halaman beranda utama yang akan ditampilkan setelah administrator berhasil melakukan proses *login*. Halaman ini memuat rangkuman navigasi sistem serta memberikan informasi ringkas terkait status aplikasi.



**Gambar 4. 7 Tampilan Layar *Dashboard***

#### 4.6.3 Upload Dataset

Halaman ini memfasilitasi administrator untuk menginisialisasi korpus data mentah ke dalam lingkungan sistem. Administrator dapat mengunggah berkas yang memuat lirik lagu R&B dan Neo-Soul dalam ekstensi berformat .csv.

**Gambar 4. 8 Tampilan Layar *Upload Dataset***

#### 4.6.4 Dataset

Halaman dataset berfungsi sebagai repositori untuk menampilkan sekumpulan dokumen lirik mentah (*raw data*) yang sebelumnya telah berhasil diunggah dan disimpan ke dalam basis data MySQL.

| ID | Judul Lagu                 | Artis     | Lirik Mentah (Snippet)                                                                 | Label Manual |
|----|----------------------------|-----------|----------------------------------------------------------------------------------------|--------------|
| 1  | Easy                       | Mac Ayres | "Oh, you heard What they say Oh, the more things change The more they stay..."         | Positif      |
| 2  | Slow Down                  | Mac Ayres | "I'm on my second guess Or maybe my second best What do I have to do To kn..."         | Netral       |
| 3  | Stay                       | Mac Ayres | "Something's in the water, something's in the rain Coulda gone forgot, but som..."     | Netral       |
| 4  | Calvin's Joint             | Mac Ayres | "I don't wanna wake up from sleeping in this same bed everyday Oh, the storm..."       | Positif      |
| 5  | Get to You Again           | Mac Ayres | "Mmm I guess we'll meet again, my friend It's not so often that you come aroun..."     | Negatif      |
| 6  | Waiting                    | Mac Ayres | "Cry out for me I hear your song Up against the breeze But moving along I take..."     | Negatif      |
| 7  | Show Me                    | Mac Ayres | "Didn't wanna start up in a losing game No, I didn't wanna feel this way I guess..."   | Positif      |
| 8  | The Devil's in the Details | Mac Ayres | "I don't wanna wake up from it You've given me a place and time They say the ..."      | Positif      |
| 9  | Change Ya Mind             | Mac Ayres | "Can't feel my fingers Still feelin' you Won't you come figure Out what I could d..."  | Negatif      |
| 10 | Roses                      | Mac Ayres | "Get this out the way, babe Tell me that you're only human I've made my mista..."      | Positif      |
| 11 | Under                      | Mac Ayres | "You're moving so close to me baby It don't feel that close enough All of this dis..." | Positif      |

Gambar 4. 9 Tampilan Layar *Dataset*

#### 4.6.5 Preprocessing

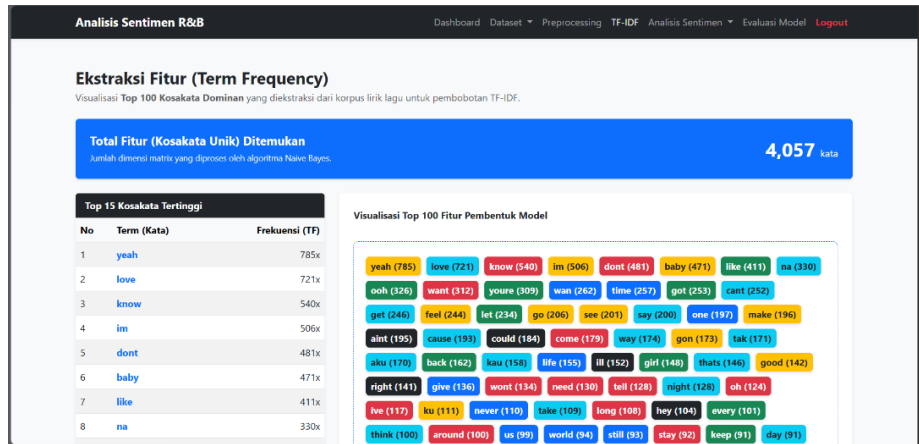
Halaman *preprocessing* menampilkan rekam jejak transformasi leksikal secara mendetail. Pada halaman ini, administrator dapat melihat perubahan bentuk lirik lagu mentah yang dieksekusi langkah demi langkah mulai dari *case folding* hingga *stemming* menjadi lirik bersih.

| No | Info Lagu                     | Lirik Mentah                          | 1. Case Folding                      | 2. Cleaning                          | 3. Tokenizing                                         | 4. Stopword Removal                 |
|----|-------------------------------|---------------------------------------|--------------------------------------|--------------------------------------|-------------------------------------------------------|-------------------------------------|
| 1  | Easy<br>Mac Ayres             | "Oh, you heard What they say Oh..."   | oh, you heard what they say oh...    | oh you heard what they say oh t...   | ['oh', 'you', 'heard', 'what', 'they', ...]           | heard say things change stay ain... |
| 2  | Slow Down<br>Mac Ayres        | "I'm on my second guess Or may..."    | i'm on my second guess or mayb...    | im on my second guess or mayb...     | ['im', 'on', 'my', 'second', 'guess', ...]            | im second guess maybe second ...    |
| 3  | Stay<br>Mac Ayres             | "Something's in the water, someth..." | something's in the water, somet...   | somethings in the water somethi...   | ['something's', 'in', 'the', 'water', 's...', ...]    | somethings water somethings rat...  |
| 4  | Calvin's Joint<br>Mac Ayres   | "I don't wanna wake up from slee..."  | i dont wanna wake up from slee...    | i dont wanna wake up from slee...    | ['I', 'dont', 'wanna', 'wake', 'up', ...]             | dont wanna wake sleeping bed ...    |
| 5  | Get to You Again<br>Mac Ayres | "Mmm I guess we'll meet again, ..."   | mmm i guess we'll meet again, ...    | mmm i guess we'll meet again m...    | ['mmm', 'i', 'guess', 'we', 'll', 'meet', ...]        | mmm guess we'll meet friend off...  |
| 6  | Waiting<br>Mac Ayres          | "Cry out for me I hear your song ..." | cry out for me i hear your song u... | cry out for me i hear your song u... | ['cry', 'out', 'for', 'me', 'i', 'hear', 'y...', ...] | cry hear song breeze moving ala...  |
| 7  | Show Me<br>Mac Ayres          | "Didn't wanna start up in a losin..." | didnt wanna start up in a losin...   | didnt wanna start up in a losin...   | ['didnt', 'wanna', 'start', 'up', 'in', ...]          | didnt wanna start losing game d...  |

Gambar 4. 10 Tampilan Layar *Preprocessing*

#### 4.6.6 TF-IDF

Halaman ini memvisualisasikan hasil ekstraksi fitur matematis. Sistem menampilkan daftar matriks pembobotan yang menunjukkan besaran nilai *Term Frequency - Inverse Document Frequency* (TF-IDF) dari setiap *token* kata yang unik.



Gambar 4. 11 Tampilan Layar *TF-IDF*

#### 4.6.7 Hasil Klasifikasi

Halaman hasil klasifikasi menyajikan tabel komparasi krusial yang menampilkan dokumen lirik bersih, label sentimen aktual (*ground truth*), bersanding dengan label sentimen hasil tebakan prediksi dari algoritma *Multinomial Naive Bayes*.

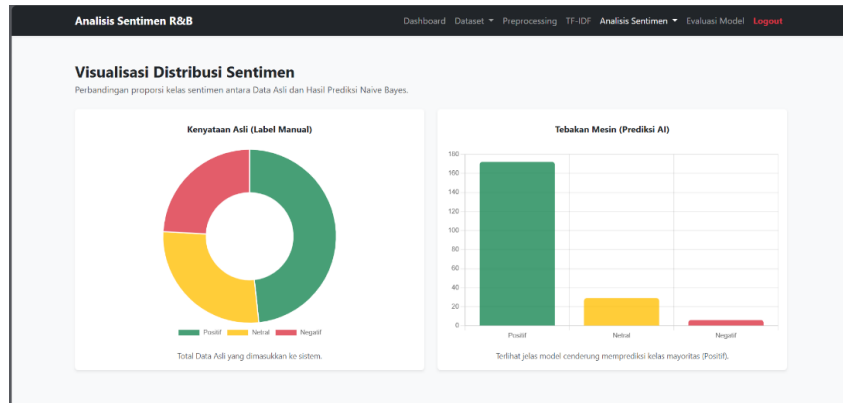
**Tabel Hasil Klasifikasi Naive Bayes**  
Perbandingan antara kenyataan label asli dan hasil tebakan sistem.

| No | Judul Lagu                 | Artis     | Kenyataan Asli | Tebakan AI | Status Analisis |
|----|----------------------------|-----------|----------------|------------|-----------------|
| 1  | Easy                       | Mac Ayres | Positif        | Positif    | Cocok           |
| 2  | Slow Down                  | Mac Ayres | Netral         | Positif    | Meleset         |
| 3  | Stay                       | Mac Ayres | Netral         | Positif    | Meleset         |
| 4  | Calvin's Joint             | Mac Ayres | Positif        | Positif    | Cocok           |
| 5  | Get to You Again           | Mac Ayres | Negatif        | Positif    | Meleset         |
| 6  | Waiting                    | Mac Ayres | Negatif        | Positif    | Meleset         |
| 7  | Show Me                    | Mac Ayres | Positif        | Positif    | Cocok           |
| 8  | The Devil's in the Details | Mac Ayres | Positif        | Positif    | Cocok           |
| 9  | Change Ya Mind             | Mac Ayres | Negatif        | Positif    | Meleset         |
| 10 | Roses                      | Mac Ayres | Positif        | Positif    | Cocok           |
| 11 | Under                      | Mac Ayres | Positif        | Positif    | Cocok           |

Gambar 4. 12 Tampilan Layar Hasil Klasifikasi

#### 4.6.8 Distribusi Sentimen

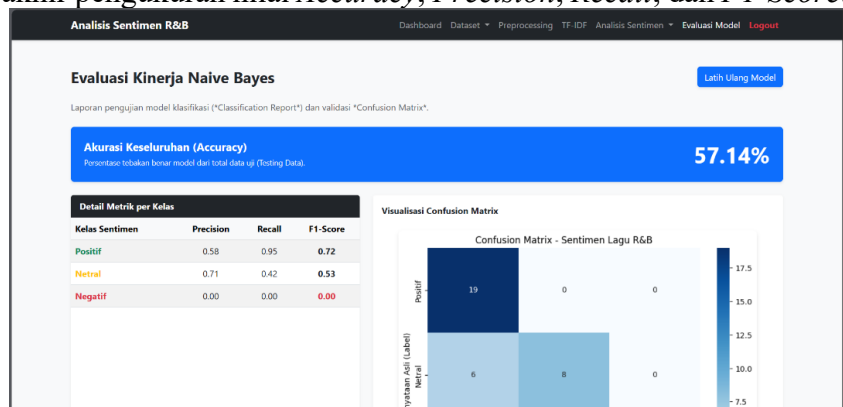
Halaman ini menyajikan representasi visual berupa grafik distribusi yang memetakan agregat total kelas sentimen (Positif, Netral, dan Negatif) dari seluruh dokumen lirik yang berhasil diprediksi oleh sistem.



Gambar 4. 13 Tampilan Layar Distribusi Sentimen

#### 4.6.9 Evaluasi Model

Halaman evaluasi model bertindak sebagai metrik penilaian akhir dari algoritma sistem. Pada halaman ini, sistem menyajikan parameter kalkulasi kinerja berupa tabel perhitungan *Confusion Matrix* beserta hasil akhir pengukuran nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.



Gambar 4. 14 Tampilan Layar Evaluasi Model

## BAB V PENUTUP

### 5.1 Kesimpulan

Berdasarkan serangkaian tahapan penelitian, implementasi sistem, hingga pengujian evaluasi yang telah dilakukan terhadap analisis sentimen lirik lagu R&B dan Neo-Soul menggunakan algoritma *Multinomial Naive Bayes*, dapat ditarik beberapa kesimpulan utama sebagai berikut:

1. Keberhasilan Implementasi Arsitektur Sistem

Sistem analisis sentimen telah berhasil dibangun dan diintegrasikan dengan modul *Text Mining*. Rangkaian proses pembersihan data teks melalui *preprocessing* (*case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*) terbukti mampu mengeliminasi *noise* leksikal. Data yang bersih tersebut juga telah sukses ditransformasikan menjadi matriks bobot matematis yang terukur menggunakan metode ekstraksi fitur *Term Frequency - Inverse Document Frequency* (TF-IDF).

2. Kinerja Algoritma Klasifikasi

Evaluasi model *Multinomial Naive Bayes* menggunakan metode *Confusion Matrix* dengan skenario *data splitting* berasio 80:20 (165 dokumen data latih dan 42 dokumen data uji) menghasilkan nilai akurasi akhir sebesar **57,14%**. Model menunjukkan sensitivitas yang sangat baik dalam mengenali lirik berkelas Positif (mencapai nilai *Recall* 0.95), namun terindikasi mengalami kesulitan yang signifikan dalam memprediksi kelas Negatif.

3. Ketimpangan Prediksi (*Skewness*)

Pemrosesan agregat terhadap total 207 dokumen korpus lirik menghasilkan ketimpangan distribusi tebakan kelas sentimen, di mana sistem secara mayoritas memprediksi 168 lagu ke dalam sentimen Positif. Hal ini membuktikan bahwa algoritma probabilistik konvensional sangat rentan terhadap bias leksikal ketika dihadapkan pada teks bernuansa seni yang sarat akan majas.

4. Analisis Anomali Leksikal Genre R&B

Berdasarkan ekstraksi kata kunci dominan, ditemukan bahwa karakteristik lirik genre R&B dan Neo-Soul memiliki tingkat ambiguitas emosional yang ekstrem. Kata-kata afeksi universal seperti "*love*", "*baby*", dan "*feel*" tetap mendominasi frekuensi kemunculan pada lirik lagu patah hati atau perpisahan (label aktual Negatif). Irisan leksikal yang identik antara lagu bernuansa bahagia dan sedih inilah yang bertindak sebagai hambatan utama, mengelabui kalkulasi probabilitas agregat mesin, dan memaksa model untuk menjatuhkan keputusan ke arah sentimen Positif secara sepihak.

## 5.2 Saran

Merujuk pada kendala dan batasan masalah yang ditemukan selama proses penelitian, terdapat beberapa rekomendasi akademis dan teknis yang dapat diajukan untuk pengembangan penelitian selanjutnya:

1. Penerapan Algoritma yang Peka Konteks (*Context-Aware*)

Mengingat keterbatasan algoritma *Multinomial Naive Bayes* dalam memahami susunan kalimat (*bag-of-words*), penelitian selanjutnya disarankan untuk menggunakan algoritma klasifikasi yang mampu memahami konteks runtutan kata, seperti *Support Vector Machine* (SVM) atau arsitektur *Deep Learning* (*Long Short-Term Memory* / BERT) untuk mengatasi ambiguitas pada karya sastra.

2. Penambahan Leksikon Khusus R&B

Disarankan untuk menambahkan tahap *preprocessing* berupa kamus standarisasi idiom dan bahasa gaul (*slang words*) bahasa Inggris yang secara khusus sering digunakan dalam kultur musik R&B dan Neo-Soul, guna menekan angka *noise* leksikal sebelum masuk ke tahap ekstraksi fitur.

3. Ekspansi Volume Dataset Latih

Disarankan untuk memperbesar volume serta menyeimbangkan porsi (*balancing data*) korpus dokumen dataset. Dengan jumlah data latih yang lebih masif, mesin diharapkan memiliki referensi kalkulasi *likelihood* yang lebih kaya dalam membedakan lagu romantis (Positif) dan lagu patah hati (Negatif) yang menggunakan struktur diksi serupa.

## DAFTAR PUSTAKA

- Hapsari, I., & Pandya Wisesa, S. (2025). Evaluasi Model Prediksi Curah Hujan Berbasis Machine Learning di Kota Bandung. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 11(2), 136–143. <https://doi.org/10.25077/teknosi.v11i2.2025.136-143>
- Hider, P., & Lee, D. (2023). A polyphony of characteristics: An analysis of the categorisation of music's subgenres. *Journal of Information Science*. <https://doi.org/10.1177/01655515231203511>
- Josen Limbong, J. A., Sembiring, I., Dwi Hartomo, K., Kristen Satya Wacana, U., & Korespondensi, P. (2022). ANALISIS KLASIFIKASI SENTIMEN ULASAN PADA E-COMMERCE SHOPEE BERBASIS WORD CLOUD DENGAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR ANALYSIS OF REVIEW SENTIMENT CLASSIFICATION ON E-COMMERCE SHOPEE WORD CLOUD BASED WITH NAÏVE BAYES AND K-NEAREST NEIGHBOR METHODS. <https://doi.org/10.25126/jtiik.202294960>
- Kholida Zia Abidin, Arief Setyanto, & Rudyanto Arief. (2023). Systematic Literature Review terhadap Klasifikasi Emosi pada Lirik Lagu Berbahasa Ambon menggunakan Metode Bidirectional LSTM dengan Glove Word Representation Weighting. *Pixel :Jurnal Ilmiah Komputer Grafis*, 16(2), 235–244. <https://doi.org/10.51903/pixel.v16i2.1641>
- Manueke, G. W., Lumando, L., & Manueke, E. G. Y. (2023). Analisis Sentimen Lirik Lagu Populer Dengan Pemanfaatan Teknik Natural Language Processing. In *Global Science* (Vol. 4, Number 1).
- Ode, P., Ardika, I., Gusti, I., Anom, N., Putra, C., Raya, J., Udayana, K., Jimbaran, B., Selatan, K., & Indonesia, B. (2022). *Klasifikasi Lirik Lagu Bertema Lingkungan dengan Metode Naive Bayes*.
- Odhianto, Y. O., Swanjaya, D., & Sahertian, J. (2025). Optimalisasi Latent Dirichlet Allocation untuk Ekstraksi Topik Utama dalam Teks Dongeng 1\* Penulis Korespondens : Yosan Okta Odhianto. In *Seminar Nasional Inovasi Teknologi* (Vol. 9).
- Ramadhan As'ari, W., Arifin, M., Fithri, D. L., & Setiaji, P. (2025). ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI SPOTIFY DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAIVE BAYES. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 9, Number 2).
- Riyandona, S. A., Rahaningsih, N., Dana, R. D., & Mulyawan, -. (2025). IMPLEMENTASI MODEL ANALISIS SENTIMEN TERHADAP GRUP MUSIK BTS MENGGUNAKAN METODE NAÏVE BAYES. *Jurnal*

- Informatika Dan Teknik Elektro Terapan*, 13(1).  
<https://doi.org/10.23960/jitet.v13i1.5816>
- Rosebaugh, C., & Shamir, L. (2022). Data Science Approach to Compare the Lyrics of Popular Music Artists. *Unisia*, 40(1), 1–26.  
<https://doi.org/10.20885/unisia.vol40.iss1.art1>
- Sabda Husada, M., & Yulianto, S. J. (2026). *Analisis Emosi dalam Lirik Lagu menggunakan Natural Language Processing* (Vol. 7, Number 1).  
<https://journal.stmiki.ac.id>
- Salsabila Nuzula, M., & Djauhari, O. S. (2025). Discourse of Apology and Plea in R&B Lyrics: A Pragmatic Analysis of “On Bended Knee.” *Jurnal Pustaka Cendekia Pendidikan*, 03(01), 113–118.
- Syafaah, M. I., & Lestandy, L. (2022). *Emotional Text Classification Using TF-IDF (Term Frequency-Inverse Document Frequency) And LSTM (Long Short-Term Memory)* (Vol. 10, Number 2).  
[https://atapdata.ai/dataset/192/HIMPUNAN\\_DATA\\_E](https://atapdata.ai/dataset/192/HIMPUNAN_DATA_E)
- Tri Widodo, F., & Setyawan, M. B. (2025). *Implementasi Text Mining Pada Analisis Sentimen Pemain Naturalisasi Timnas Indonesia Dengan Algoritma Naïve Bayes dan Support Vector Machine (SVM)*.
- Virginia, C., & Irsyad, H. (2025). REKOMENDASI LAGU BERDASARKAN JENIS MOOD MENGGUNAKAN TF-IDF DAN COSINE SIMILARITY. *Jurnal Insand Comtech*, 10(1).

## LAMPIRAN

### 1. Surat Validasi Pakar

#### **SURAT VALIDASI PAKAR LABEL DATA PENELITIAN TUGAS AKHIR**

##### **IDENTITAS VALIDATOR:**

Nama : Nadhif Rafii Muhamad, S.Sn.  
Bidang Keahlian : Seni Musik  
Asal Instansi : Institut Kesenian Jakarta (IKJ)  
Tanggal Validasi : 30 Juni 2026

Dengan ini menyatakan bahwa saya telah melakukan peninjauan dan verifikasi terhadap hasil pelabelan data pada dataset lirik lagu yang digunakan dalam penelitian tugas akhir mahasiswa berikut:

Nama : Muhammad Andika Anggara Lukman Saudah  
NIM : 2211501248  
Program Studi : Teknik Informatika  
Topik Tugas Akhir : ANALISIS SENTIMEN PADA LIRIK LAGU R&B DAN NEO-SOUL MENGGUNAKAN TF-IDF DAN ALGORITMA NAIVE BAYES  
Dataset : [Genius.com] (Terdiri dari 207 lirik lagu)

Adapun verifikasi dilakukan terhadap dataset lirik lagu genre R&B dan Neo-Soul yang telah diberi label sentimen (Positif, Netral, dan Negatif). Berdasarkan peninjauan saya terhadap representasi makna, gaya bahasa/metafora dalam lirik, dan kesesuaian label dengan isi pesan lagu tersebut, saya menyatakan bahwa dataset telah diberi label secara objektif, sesuai, dan layak digunakan untuk keperluan analisis sentimen.

Demikian lembar verifikasi ini dibuat untuk dapat digunakan sebagaimana mestinya.

Jakarta, 30 Juni 2026



**Nadhif Rafii Muhamad, S. Sn.**

## 2. Berita Acara

8/4/26, 3:23 PM

Universitas Budi Luhur & Akademi Sekretari Budi Luhur



### **BERITA ACARA SIDANG PENDADARAN TUGAS AKHIR**

S/UBL/FTI/0203/VII/26

Pada hari ini, Senin 13 Juli 2026 telah dilaksanakan Ujian Sidang Pendadaran Tugas Akhir sebagai berikut:

Judul : ANALISIS SENTIMEN PADA LIRIK LAGU R&B DAN NEOSOUL MENGGUNAKAN TF-IDF DAN ALGORITMA NAIVE BAYES

Nama : Muhammad Andika Anggara Lukman Saudah

NIM : 2211501248

Dosen Pembimbing : Subandi, S.Kom., M.Kom.

Berdasarkan penilaian pada Presentasi + Demo, Penulisan, Penguasaan Materi, Penguasaan Program maka Mahasiswa tersebut di atas dinyatakan:

#### **LULUS**

dengan nilai angka: 80 huruf: A-

Mahasiswa tersebut di atas wajib menyerahkan hasil perbaikan tulisan Tugas Akhir dalam bentuk terjilid sesuai dengan Panduan Perbaikan Tugas Akhir, selambat-lambatnya Senin 27 Juli 2026.

#### Panitia Penguji:

1. Ketua Dewi Kusumaningsih, S.Kom., M.Kom.
2. Anggota Mufti, S.T., M.Kom.
3. Moderator Subandi, S.Kom., M.Kom.

#### Keterangan:

Nilai Huruf: A:85-100 A-:80-84,99 B+:75-79,99 B:70-74,99 B-:65-69,99 C:60-64,99 D:40-59,99 E-:0-39,99