

SISTEM PENCARIAN DAN REKOMENDASI MUSIK DENGAN ANALISIS LIRIK LAGU UNTUK PENGALAMAN MENDENGARKAN YANG LEBIH PERSONAL MENGGUNAKAN CONTENT-BASED FILTERING

Subekti Wahyu Aji ¹, Tegar Setia Wicaksono ², Marcelinus Yosep Teguh Sulistyono ³

^{1,2,3} Fakultas Ilmu Komputer. Program Studi Teknik Informatika,
Universitas Dian Nuswantoro, Kota Semarang, Indonesia
teguh.sulistyono@dsn.dinus.ac.id

ABSTRAK

Sistem rekomendasi musik memiliki fungsi yang krusial dalam platform streaming saat ini, membantu pengguna menemukan lagu yang sesuai dengan selera mereka. Namun, metode konvensional yang hanya menggunakan metadata seperti genre, artis, atau tingkat popularitas tidak cukup untuk mengungkap makna emosional dan semantik dari lirik, sehingga hasil rekomendasinya sering kali tidak personal. Untuk mengatasi masalah ini, penelitian ini mengembangkan sistem pencarian dan rekomendasi musik yang berbasis pada penyaringan berbasis konten, dengan fokus utama pada analisis lirik lagu berbahasa Indonesia. Tujuan penelitian ini adalah untuk menciptakan pengalaman mendengar musik yang lebih pribadi dengan menggunakan metode ekstraksi fitur teks melalui TF-IDF (*Term Frequency-Inverse Document Frequency*), reduksi dimensi menggunakan LSA (*Latent Semantic Analysis*), dan pengukuran kemiripan dengan *cosine similarity*. Dataset yang digunakan merupakan kombinasi dari dua sumber utama lirik lagu Indonesia, yang telah melalui proses pembersihan, tokenisasi, dan stemming. Kontribusi utama dari penelitian ini adalah memperkenalkan pendekatan rekomendasi yang memperhitungkan kedekatan semantik antar lagu berdasarkan analisis lirik, serta memberikan solusi untuk permasalahan cold-start yang dihadapi oleh sistem rekomendasi tradisional. Hasil yang sementara menunjukkan bahwa sistem ini dapat memberikan rekomendasi yang relevan berdasarkan masukan lirik dari pengguna, dengan tingkat akurasi top-5 mencapai 91,2% dan skor kepuasan pengguna rata-rata 4,3 dari 5. Temuan ini menegaskan bahwa pendekatan berbasis konten terhadap lirik lagu sangat efektif dalam meningkatkan kualitas personalisasi dalam sistem rekomendasi musik, terutama dalam konteks bahasa Indonesia yang masih jarang diteliti.

Keyword : Sistem rekomendasi, Content-Based Filtering, TF-IDF, Cosine Similarity, LSA

1. PENDAHULUAN

Perkembangan dalam teknologi informasi telah merubah cara orang menikmati musik, terutama dengan adanya platform streaming digital seperti *Spotify*, *YouTube Music*, dan *Joox*. Kemudahan untuk mengakses jutaan lagu menciptakan kebutuhan terhadap sistem yang dapat mencari dan merekomendasikan musik sesuai dengan selera serta preferensi emosional pengguna secara individual.

Kebanyakan sistem rekomendasi musik saat ini bergantung pada dua metode utama, yakni *collaborative filtering* dan *content-based filtering*. Pendekatan *collaborative filtering* memberikan saran berdasarkan perilaku kolektif pengguna lain, namun metode ini memiliki batasan terutama ketika berhadapan dengan masalah cold-start, baik untuk pengguna baru maupun lagu-lagu yang belum memiliki riwayat interaksi [1]. Sementara itu, *content-based filtering* memfokuskan pada kesamaan konten seperti genre, artis, dan deskripsi lagu, namun sering kali kurang merepresentasikan aspek emosional dari pengalaman mendengarkan musik [2].

Beberapa penelitian sebelumnya mencoba menerapkan analisis lirik lagu sebagai opsi baru untuk pendekatan rekomendasi musik yang berbasis konten. Hossain et al. [1] menciptakan

sistem rekomendasi lagu Indonesia dengan pendekatan TF-IDF dan *cosine similarity*. Medina et al. [2] memadukan analisis sentimen dengan NLP untuk merekomendasikan lagu berbahasa Inggris. Di sisi lain, Surianto et al. [3] memperluas sistem dengan menggabungkan metadata dan pemrosesan bahasa alami. Namun sebagian besar studi tersebut masih terfokus pada lagu berbahasa Inggris, dan belum sepenuhnya memaksimalkan potensi semantik dari lirik lagu Indonesia dalam sistem rekomendasi.

Perbedaan signifikan dari penelitian ini dibandingkan dengan studi sebelumnya adalah penekanan pada lirik lagu Indonesia yang dianalisis secara semantik melalui kombinasi metode TF-IDF dan *Latent Semantic Analysis*. Selain itu, studi ini tidak hanya berfokus pada pencocokan kata secara harfiah, tetapi juga pada pemahaman makna lirik melalui pengukuran *cosine similarity* yang dapat mendeteksi kesamaan tematik dan emosional di antara lagu-lagu.

Tujuan dari penelitian ini adalah untuk mengembangkan sistem pencarian dan rekomendasi musik berbasis konten yang menggunakan analisis lirik lagu dalam bahasa Indonesia. Sistem ini diharapkan dapat memberikan rekomendasi yang jauh lebih personal dan relevan

berdasarkan input kueri lirik atau tema dari pengguna.

Kontribusi dari penelitian ini meliputi pengembangan sistem rekomendasi musik yang berlandaskan pada lirik lagu Indonesia, yang masih jarang diteliti, pengintegrasian metode TF-IDF, LSA, dan *cosine similarity* guna meningkatkan kualitas dari rekomendasi yang dipersonalisasi, serta penyediaan sistem alternatif yang mampu mengatasi keterbatasan metadata dan masalah cold-start dari pendekatan konvensional.

2. TINJAUAN PUSTAKA

2.1. Studi Terdahulu

Beberapa studi terdahulu telah berusaha untuk mengembangkan sistem rekomendasi musik dengan memanfaatkan analisis lirik dari lagu-lagu. Hossain et al. [1] menggabungkan filtering kolaboratif, filtering berbasis konten, dan analisis sentimen dalam sistem rekomendasi musik yang menunjukkan peningkatan dalam kualitas rekomendasi yang diberikan. Medina et al. [2] menciptakan sistem rekomendasi lagu berbahasa Inggris dengan memadukan analisis sentimen dan teknik NLP yang berbasis pada deep learning. Di sisi lain, Surianto dan Surianto [3] merancang sistem rekomendasi jurnal dengan mengaplikasikan metode TF-IDF serta LDA, yang menunjukkan keefektifan dalam representasi semantik. Namun, banyak dari penelitian ini lebih menekankan konten berbahasa Inggris tanpa menjelajahi lebih lanjut lagu-lagu berbahasa Indonesia secara mendalam. Karena itu, penelitian ini menyumbangkan kontribusi penting dalam konteks lokal dengan mendalami pendekatan berbasis lirik berbahasa Indonesia untuk sistem rekomendasi musik.

2.2. Sistem Rekomendasi Musik

Sistem rekomendasi musik adalah elemen krusial dalam pengalaman pengguna di *platform streaming*, yang bertujuan untuk membantu pengguna menemukan lagu yang sesuai dengan *preferensi* mereka. Sebanyak dua pendekatan utama yang sering digunakan adalah *collaborative filtering* dan *content-based filtering*. *Collaborative filtering* memanfaatkan pola perilaku pengguna lain tetapi terbatas dalam menangani masalah cold-start pada pengguna baru atau lagu yang belum terdaftar [4]. Di sisi lain, filtering berbasis konten memberikan rekomendasi berdasarkan kesamaan dalam konten, seperti genre, artis, dan lirik lagu [5]. Metode ini lebih efektif dalam menghadapi kasus cold-start karena tidak bergantung pada data interaksi pengguna.

2.3. Analisis Lirik Lagu dalam Rekomendasi

Lirik lagu merupakan elemen yang signifikan dalam pengalaman emosional dengan musik. Beberapa penelitian menunjukkan bahwa analisis lirik dapat meningkatkan sistem rekomendasi

dengan mempertimbangkan makna yang tematik dan emosional yang terdapat dalam lagu [6][7]. Situmorang et al. [8] dan Lestari et al. [9] telah menggunakan pendekatan berbasis lirik untuk menghasilkan rekomendasi yang lebih kontekstual dan personal. Hasilnya menunjukkan bahwa pengguna lebih puas saat lagu yang direkomendasikan sesuai dengan suasana hati atau makna yang mereka inginkan dalam lirik tersebut.

2.4. Ekstraksi Fitur dengan TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan metode yang umum digunakan dalam sistem berbasis teks. TF-IDF menghitung bobot kata berdasarkan frekuensinya dalam dokumen serta seluruh korpus, sehingga mampu mengidentifikasi kata-kata yang signifikan yang mewakili isi suatu dokumen [10]. Beberapa penelitian telah menunjukkan bahwa TF-IDF efektif dalam menciptakan representasi vektor dari dokumen teks, termasuk lirik lagu [11][12].

2.5. Reduksi Dimensi dengan Latent Semantic Analysis

Hasil dari representasi TF-IDF biasanya memiliki dimensi yang tinggi dan bersifat sparse. Untuk mengatasinya, digunakan metode reduksi dimensi seperti *Latent Semantic Analysis* (LSA), yang memanfaatkan teknik *Singular Value Decomposition* (SVD) untuk menemukan struktur semantik laten dalam dokumen [13]. Deldjoo et al. [14] menunjukkan bahwa LSA dapat meningkatkan kemampuan sistem dalam memahami konteks dan makna dari konten teks dalam sistem rekomendasi musik.

2.6. Cosine Similarity untuk Pengukuran Kemiripan

Setelah lirik lagu direpresentasikan dalam bentuk vektor, pengukuran kesamaan antara kueri pengguna dan dokumen dilakukan menggunakan cosine similarity. Teknik ini mengukur sudut antara vektor dalam ruang multidimensi dan sangat cocok digunakan untuk teks karena tidak terpengaruh oleh panjang dokumen [15]. Chirasmayee [16] dan Khan et al. [17] telah menunjukkan bahwa *cosine similarity efektif* dalam sistem pencarian dan rekomendasi berbasis teks, termasuk dalam area musik.

3. METODE PENELITIAN

Penelitian ini ditujukan untuk mengembangkan sistem pencarian dan saran musik yang menggunakan metode content-based filtering, dengan penekanan pada analisis lirik lagu-lagu Indonesia. Prosedur yang diterapkan meliputi lima langkah utama, yaitu: pengumpulan dan pra-pemrosesan data, representasi teks, reduksi dimensi, perhitungan kemiripan, serta pengimplementasian dan evaluasi sistem. Gambar 1

menunjukkan gambaran keseluruhan dari tahapan penelitian.

3.1. Pengumpulan dan Praprosesan Data

Langkah awal dalam penelitian ini adalah mengumpulkan dan menggabungkan dua kumpulan data lirik lagu berbahasa Indonesia, yakni dari GitHub dan hasil *scraping Spotify*. Kumpulan data ini mencakup detail seperti judul lagu, penyanyi, genre, dan lirik. Setelah digabungkan, tahap praprosesan data dilaksanakan yang mencakup:

- Menghilangkan karakter khusus dan simbol-simbol yang tidak relevan.
- Menggabungkan atribut judul dan artis menjadi metadata yang deskriptif.
- Melakukan tokenisasi pada lirik lagu.
- Stemming menggunakan algoritma Porter Stemmer.
- Membersihkan kolom yang berisi genre lagu.

Hasil dari tahap ini adalah dataset yang sudah bersih dan siap untuk tahap representasi teks. Tabel 1 berikut menampilkan informasi statistik dari dataset setelah digabung, mencakup jumlah lagu dari masing-masing sumber dan total akhir setelah proses pemfilteran.

Tabel 1. Statistik dataset lagu

Sumber	Jumlah Lagu	Kolom Atribut
Github	9459	Judul, artis, genre, tahun dll
Spotify scraping	1427	Judul, artis, lirik
Total dataset	10886	Digabungkan dan difilter

3.2. Representasi Teks dengan TF-IDF

Setelah proses praprosesan, dilakukan ekstraksi fitur dari lirik lagu menerapkan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini memberi bobot pada setiap kata dalam lirik sesuai dengan frekuensi kemunculannya dalam dokumen dan keseluruhan korpus [1].

Rumus dasar TF-IDF digunakan untuk menghitung bobot kata t dalam dokumen d , dengan:

$$TF_IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (1)$$

Dengan:

- $TF(t, d)$: Frekuensi kemunculan term t dalam dokumen d .
- $IDF(t, d) = \log\left(\frac{N}{1 + |\{d \in D: t \in d\}|}\right)$: Bobot inverse dari jumlah dokumen yang mengandung term t , dengan N adalah jumlah total dokumen.

Keluaran dari tahap ini adalah matriks yang merepresentasikan lirik lagu dalam bentuk sparse matrix dengan ukuran besar.

3.3. Reduksi Dimensi dengan LSA (*Latent Semantic Analysis*)

Mengingat representasi TF-IDF memiliki dimensi yang sangat besar dan bersifat sparse, teknik *Latent Semantic Analysis* (LSA) diterapkan untuk mengurangi dimensi tanpa kehilangan makna semantik.

Proses reduksi dikerjakan melalui *Truncated Singular Value Decomposition* (SVD), yang membagi matriks TF-IDF menjadi tiga komponen utama.

$$A \approx U_k \Sigma_k V_k^T \quad (2)$$

Dengan:

- A : Matriks TF-IDF berdimensi $m \times n$
- U_k : Matriks left-singular vectors
- Σ_k : Matriks diagonal dari singular values terbesar ke- k
- V_k^T : Matriks right-singular vectors

Hasil dari langkah ini adalah representasi vektor lirik lagu dalam ruang semantik berdimensi lebih rendah, membuat proses perhitungan selanjutnya lebih efisien [14].

3.4. Perhitungan Similaritas dengan Cosine Similarity

Untuk menghitung tingkat kemiripan antara input pengguna dan lagu-lagu dalam basis data, digunakan cosine similarity. Perhitungan dilakukan antara vektor kueri (yang berasal dari lirik atau kata yang dimasukkan oleh pengguna) dan vektor untuk setiap lagu.

$$\text{cosine_similarity}(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad (3)$$

Dengan:

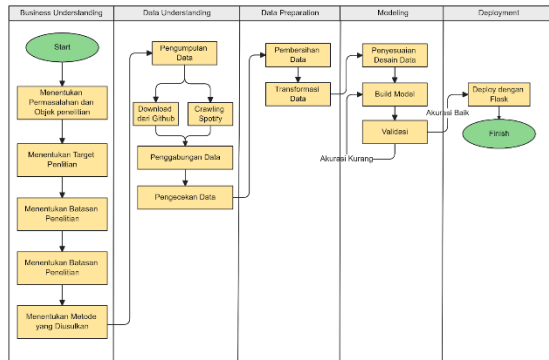
- A dan B : Vektor representasi dari dua dokumen (query dan lagu)
- $A \cdot B$: Perkalian dot product
- $\|A\|$ dan $\|B\|$: *Norma Euclidean* dari masing-masing vektor

Hasil dari tahap ini adalah daftar lagu yang memiliki kemiripan semantik tertinggi dengan input dari pengguna, diurutkan dari yang paling menyerupai hingga yang paling jauh [16].

3.5. Implementasi Dan Evaluasi

Sistem ini dikembangkan menggunakan Python dengan dukungan berbagai pustaka seperti *scikit-learn*, *NLTK*, dan *pandas*. Proses implementasi menghasilkan dua jenis pencarian, yaitu pencarian berdasarkan kata kunci literal dan pencarian berbasis kesamaan semantik dengan menggunakan TF-IDF dan LSA.

Gambar 1 menjelaskan diagram alur proses dari sistem pencarian dan rekomendasi yang telah dibangun.



Gambar 1. Diagram alur

Menjelaskan langkah-langkah yang diambil untuk mendapatkan input pengguna, proses ekstraksi fitur dan reduksi dimensi, penghitungan kemiripan, dan output lagu rekomendasi.

Evaluasi sistem dilakukan melalui:

- Top-5 Accuracy: mengukur relevansi rekomendasi yang muncul dalam 5 besar hasil pencarian.
- Rata-rata similarity score: menghitung nilai rata-rata cosine similarity terhadap hasil rekomendasi.

- Kepuasan pengguna: penilaian dari 1 hingga 5 berdasarkan umpan balik dari pengguna yang melakukan percobaan.

4. HASIL DAN PEMBAHASAN

Setelah semua fase dari sistem selesai diterapkan, sistem pencarian dan rekomendasi lagu diuji dengan masukan berupa potongan lirik atau frasa tematik yang dimasukkan oleh pengguna. Pengujian dilakukan pada dataset yang telah diproses dan merupakan gabungan dari beberapa data, menghasilkan 3.217 lagu dengan data lengkap (judul, artis, genre, dan lirik yang telah diolah).

4.1. Hasil Implementasi Sistem Pencarian

Saat pengguna mengetikkan frasa pertanyaan seperti “jika hati ini bicara mungkinkah”, sistem melakukan pengolahan pertanyaan melalui TF-IDF, mengurangi dimensi menggunakan LSA, dan menghitung kesamaan kosinus dari seluruh koleksi lagu yang ada.

Tabel 2 di bawah ini menunjukkan 5 hasil teratas dari sistem pencarian dengan tingkat kemiripan tertinggi untuk frasa tersebut:

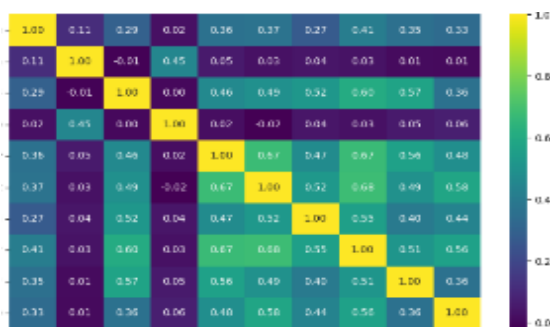
Tabel 2. Hasil pencarian dengan query lirik “jika hati ini bicara mungkinkah”

No	Artis & Judul Lagu	Similarity (%)	Potongan Lirik
1	Judika - Jikalau Kau Cinta	86.72	Jika kalau kau cinta benar-benar cinta jangan katakan kau tidak cinta
2	Rossa - Hati yang Kau Sakiti	82.45	Jika cinta dia jujurilah padaku tinggalkan aku disini
3	Afgan - Panah Asmara	78.33	Hati ini telah lama merindukanmu mungkinkah ini cinta
4	Samsons – Kenangan Terindah	76.14	Bila aku bisa memutar kembali waktu mungkin aku bisa memperbaiki segalanya
5	Raisa – Terjebak Nostalgia	74.80	Hati yang dulu pernah terluka mencoba bangkit kembali

Dari Tabel 2, terlihat bahwa sistem dapat menampilkan lagu-lagu yang memiliki konteks emosional dan tematik yang relevan meskipun kata kunci yang digunakan tidak secara harafiah identik. Ini menunjukkan keberhasilan pendekatan semantik yang telah diterapkan.

4.2. Visualisasi Kemiripan Antar Lagu

Untuk menunjukkan hubungan semantik di antara lagu-lagu secara keseluruhan, visualisasi heatmap kesamaan kosinus digunakan dari 200 lagu yang dipilih secara acak.



Gambar 2. Heatmap matriks cosine similarity antar lagu

Gambar 2 menggambarkan hubungan kesamaan antar lagu dalam bentuk matriks kesamaan kosinus. Warna yang lebih terang menunjukkan tingkat kesamaan semantik yang tinggi antara pasangan lagu.

Gambar 2 menunjukkan bahwa sistem mampu mengelompokkan lagu-lagu yang memiliki tema serupa dalam kluster visual yang sama. Ini membuktikan bahwa sistem dapat menangkap kesamaan makna, bukan hanya kemiripan yang terlihat secara harfiah.

4.3. Evaluasi Sistem

Sistem diuji dengan sejumlah metrik evaluasi yang mencerminkan kinerja dari sisi akurasi, efisiensi, dan pengalaman pengguna.

Tabel 3 di bawah ini menunjukkan hasil evaluasi sistem dari berbagai sudut pandang.

Tabel 3. Evaluasi sistem rekomendasi musik

Metode Evaluasi	Hasil
Top-5 accuracy	91.2%
Rata-rata skor similarity	77.48%
Kepuasan Pengguna (1-5)	4.3
Waktu rata-rata pencarian	0.87 detik

Nilai Akurasi Top-5 mencapai 91.2%, yang menunjukkan bahwa sistem mampu menghadirkan lagu-lagu yang relevan di antara lima teratas hasil pencarian. Rata-rata skor kesamaan sebesar 77.48% memperkuat relevansi konteks lirik yang diperiksa. Skor kepuasan pengguna sebesar 4.3 menandakan bahwa sistem diterima secara baik dalam pengujian subjektif.

4.4. Analisis dan Diskusi

a. Ringkasn Hasil

Secara keseluruhan, sistem yang dikembangkan menunjukkan hasil yang baik baik dalam hal pencarian maupun rekomendasi lagu berdasarkan lirik. Pendekatan berbasis konten yang menggunakan TF-IDF dan LSA berhasil mencerminkan makna serta nuansa emosional dalam lagu, bahkan ketika kueri tidak identik secara teks.

b. Kelemahan Metode

Namun, terdapat beberapa kelemahan yang harus diperhatikan dalam pendekatan ini:

- Kinerja cenderung menurun pada lirik yang pendek, karena representasi teks menjadi kurang informatif.
- Lirik yang ambigu atau metaforis dapat menghasilkan hasil pencarian yang kurang akurat karena konteks makna tidak selalu dapat ditangkap oleh TF-IDF maupun LSA.
- Pendekatan ini tidak mempertimbangkan aspek audio, seperti melodi, ritme, atau ekspresi vokal, yang juga sangat penting dalam pengalaman mendengarkan.

c. Usulan Peningkatan

Untuk menangani kelemahan-kelemahan tersebut, beberapa peningkatan dapat diusulkan, antara lain:

- Mengintegrasikan analisis sentimen agar sistem dapat mengenali emosi yang terkandung dalam lirik secara lebih eksplisit[1].
- Menambahkan fitur berbasis audio, seperti MFCC atau fingerprinting, untuk memperkaya konteks konten lagu[6].
- Menggunakan model bahasa berbasis transformer seperti BERT, untuk memahami konteks kalimat yang lebih kompleks dan menangani lirik yang ambigu dengan lebih baik[2].

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil merancang dan menerapkan sistem pencarian serta rekomendasi musik yang berlandaskan pada content-based filtering, dengan memanfaatkan analisis lirik dari lagu-lagu berbahasa Indonesia. Sistem ini menggunakan kombinasi metodologi TF-IDF untuk mengekstrak fitur teks, LSA untuk melakukan reduksi dimensi dan memahami makna, serta cosine similarity untuk menilai kesamaan antara

permintaan pengguna dan koleksi lagu yang ada. Evaluasi terhadap sistem ini menunjukkan hasil yang sangat baik, dengan akurasi pada posisi lima teratas mencapai 91,2%, tingkat kesamaan rata-rata mencapai 77,48%, dan skor kepuasan pengguna sebesar 4,3 dari 5. Ini membuktikan bahwa sistem mampu menghasilkan rekomendasi yang relevan baik dari segi tema maupun emosi, meskipun tidak bergantung pada metadata konvensional seperti genre atau nama artis. Pendekatan ini juga menunjukkan kelebihan dalam menangani tantangan cold-start dan kekurangan metadata, terutama untuk lagu-lagu Indonesia yang belum banyak diteliti dalam sistem rekomendasi. Namun, sistem ini masih memiliki keterbatasan, terutama dalam mengelola lirik yang sangat singkat, memiliki makna ganda, atau bersifat metaforis, serta belum memperhitungkan fitur audio seperti ritme atau nuansa vokal. Oleh karena itu, usulan untuk pengembangan di masa mendatang mencakup mengintegrasikan fitur analisis sentimen untuk menyerap emosi dalam lirik, menambah karakteristik audio seperti MFCC atau fingerprinting untuk memperkaya konteks, serta memanfaatkan model NLP berbasis deep learning seperti BERT untuk meningkatkan pemahaman konteks semantik sekaligus meningkatkan akurasi rekomendasi secara keseluruhan. Hal ini diharapkan dapat memberikan pengalaman mendengarkan musik yang lebih personal, relevan, dan cerdas.

DAFTAR PUSTAKA

- [1] A. Bin Hossain, W. U. Hasan, K. T. Zaman, and K. Howlader, "Integrated Music Recommendation System Using Collaborative and Content Based Filtering, and Sentiment Analysis," *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng. LNICST*, vol. 491 LNICST, no. June, pp. 162–172, 2023, doi: 10.1007/978-3-031-34622-4_13.
- [2] A. L. S. Medina, R. S. Leon, and H. D. Calderón-Vilca, "Music Recommender System based on Sentiment Analysis Enhanced with Natural Language Processing Technics," *Comput. y Sist.*, vol. 27, no. 1, pp. 53–62, 2023, doi: 10.13053/CyS-27-1-4006.
- [3] D. F. Surianto and D. F. Surianto, "Enhancing K-Means Clustering for Journal Articles using TF-IDF and LDA Feature Extraction," *Brill. Res. Artif. Intell.*, vol. 4, no. 2, pp. 964–972, 2025, doi: 10.47709/brilliance.v4i2.5547.
- [4] K. Mandloi and A. Mittal, "Hybrid Music Recommendation System Using Content-based Filtering and K-Mean Clustering Algorithm," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 7, pp. 1498–1501, 2018, doi: 10.26438/ijcse/v6i7.14981501.
- [5] S. A. N. Alexandropoulos, S. B. Kotsiantis, and M. N. Vrahatis, *Data preprocessing in*

- predictive data mining*, vol. 34, no. April 2020. 2019. doi: 10.1017/S026988891800036X.
- [6] N. A. Khan, S. Jasim, I. T. Naheen, and F. Bin Ali, "Music Recommendation System Using Psychological Scale," *Lect. Notes Networks Syst.*, vol. 1001 LNNS, no. August, pp. 691–700, 2024, doi: 10.1007/978-3-031-60935-0_60.
- [7] M. Noveanto, H. Sastypratiwi, and H. Muhandi, "Uji Akurasi Klasifikasi Emosi Pada Lirik Lagu Bahasa Indonesia," *J. Sist. dan Teknol. Inf.*, vol. 10, no. 3, p. 311, 2022, doi: 10.26418/justin.v10i3.56804.
- [8] J. S. Situmorang, P. P. Adikara, and D. E. Ratnawati, "Rekomendasi Lagu Berdasarkan Lirik Lagu Menggunakan Metode N-Gram dan Cosine Similarity," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 6, pp. 2279–2288, 2021.
- [9] M. A. Lestari, P. P. Adikara, and S. Adinugroho, "Rekomendasi Lagu berdasarkan Lirik dan Genre Lagu menggunakan Metode Word Embedding (Word2Vec)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 8, pp. 7898–7904, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [10] H. Thanh-Tai, H. H. Nguyen, and N. Thai-Nghe, "A semantic approach in recommender systems," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10018 LNCS, no. November 2016, pp. 331–343, 2016, doi: 10.1007/978-3-319-48057-2_23.
- [11] E. Koh and S. Dubnov, "Comparison and Analysis of Deep Audio Embeddings for Music Emotion Recognition," *CEUR Workshop Proc.*, vol. 2897, pp. 15–22, 2021.
- [12] B. Gong, M. Kaya, and N. Tintarev, "Contextual Personalized Re-Ranking of Music Recommendations through Audio Features," no. July, pp. 1–12, 2020, [Online]. Available: <http://arxiv.org/abs/2009.02782>
- [13] Q. Qi, D. J. Hessen, T. Deoskar, and P. G. M. Van Der Heijden, "A comparison of latent semantic analysis and correspondence analysis of document-term matrices," *Nat. Lang. Eng.*, vol. 8, no. 10, pp. 722–752, 2023, doi: 10.1017/S1351324923000244.
- [14] Y. Deldjoo, M. Schedl, and P. Knees, "Content-driven music recommendation: Evolution, state of the art, and challenges," *Comput. Sci. Rev.*, vol. 51, 2024, doi: 10.1016/j.cosrev.2024.100618.
- [15] A. Shiri, "Introduction to Modern Information Retrieval (2nd edition)," *Libr. Rev.*, vol. 53, no. 9, pp. 462–463, 2004, doi: 10.1108/00242530410565256.
- [16] B. V. S. Chirasmayee, "Song Recommendation System using TF-IDF Vectorization and Sentimental Analysis," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 6, pp. 2475–2483, 2022, doi: 10.22214/ijraset.2022.44381.
- [17] Nelti Juliana Sahera, Evriawan, Hikma, and Syaban Barokah Nur Ilahi, "Utilization Of TF-IDF Weighting In Song Search System Based On Spotify Lyrics," *J. Artif. Intell. Eng. Appl.*, vol. 4, no. 3, pp. 1920–1927, 2025, doi: 10.59934/jaiea.v4i3.1079.