Jurnal Serina Sains, Teknik dan Kedokteran
Vol. 01, No. 02, Sep 2023: hlm 455-462 ISSN-L 2986-5204 (Versi Elektronik)

PENERAPAN METODE CONTENT BASED FILTERING DALAM IMPLEMENTASI SISTEM REKOMENDASI MUSIK

Ghatra Vaganzha Rihi Pake[1], Viny Christanti Mawardi[2], Tri Sutrisno[3]

[1]Fakultas Teknik Informasi, Universitas Tarumanagara Jakarta
Email: ghatra.535180142@stu.untar.ac.id
[2] Fakultas Teknik Informasi, Universitas Tarumanagara Jakarta
Email: vinym@fti.untar.ac.id
[3]Fakultas Teknik Informasi, Universitas Tarumanagara Jakarta
Email: tris@fti.untar.ac.id

Masuk : 26-09-2023, revisi: 18-10-2023, diterima untuk diterbitkan : 25-10-2023

ABSTRAK

Terdapat perubahan signifikan dalam cara manusia mengakses musik berkat perkembangan teknologi informasi dan popularitas platform streaming musik seperti Spotify, Apple Music, dan Deezer. Meskipun platform-platform ini memberikan akses tak terbatas ke berbagai jenis musik, pengguna sering menghadapi kesulitan dalam menemukan lagu yang sesuai dengan preferensi mereka. Dalam konteks ini, sistem rekomendasi musik berbasis konten menjadi krusial. Salah satu metode yang digunakan dalam pengembangan sistem ini adalah "Content-Based Filtering," yang berfokus pada analisis karakteristik lagu seperti genre, tempo, instrumen, lirik, dan menggunakan perhitungan cosine similarity untuk memberikan rekomendasi yang sesuai. Perancangan ini bertujuan untuk mengimplementasikan algoritma yang ada dalam sistem rekomendasi musik berbasis konten dengan tujuan memahami kinerja mereka dalam memberikan rekomendasi yang akurat. Hasil dari perancangan ini diharapkan dapat memberikan panduan untuk pengembangan sistem rekomendasi musik yang lebih baik, meningkatkan pengalaman musik pengguna di platform-platform musik digital.

Kata Kunci: Teknologi Informasi; Sistem Rekomendasi; Content-Based Filtering; Cosine Similarity

ABSTRACT

There has been a significant shift in how people access music due to the advancements in information technology and the popularity of music streaming platforms such as Spotify, Apple Music, and Deezer. While these platforms provide unlimited access to various types of music, users often face challenges in finding songs that align with their preferences. In this context, content-based music recommendation systems have become crucial. One of the methods used in developing these systems is "Content-Based Filtering," which focuses on analyzing song characteristics such as genre, tempo, instruments, lyrics, and utilizes cosine similarity calculations to provide relevant recommendations. This study aims to implement existing algorithms within content-based music recommendation systems with the goal of understanding their performance in delivering accurate recommendations. The results of this study are expected to offer insights for the development of more effective music recommendation systems, enhancing the music listening experience for users on digital music platforms.

Keywords: *Information Technology; Recommendation System; Content-Based Filtering; Cosine Similarity*

1. PENDAHULUAN

Perkembangan teknologi informasi dan popularitas platform streaming musik telah mengubah cara manusia mengakses dan menikmati musik. Penggunaan platform musik digital yang semakin meluas, seperti Spotify, Apple Music, dan Deezer, telah menyediakan akses tak terbatas ke berbagai genre dan artis musik di seluruh dunia [1]. Namun, dengan begitu banyaknya pilihan musik yang tersedia, pengguna seringkali menghadapi kesulitan dalam menemukan lagu-lagu yang sesuai dengan preferensi musik mereka. Dalam konteks ini, sistem rekomendasi musik

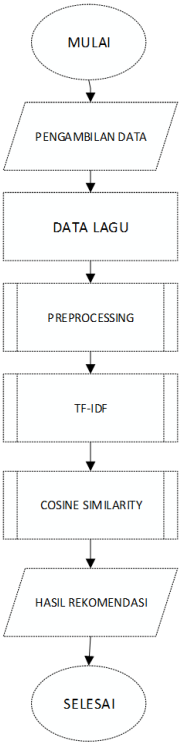
menjadi penting untuk membantu pengguna menemukan dan menikmati musik yang cocok dengan selera musik pengguna.

Salah satu pendekatan yang digunakan dalam pengembangan sistem rekomendasi musik adalah "Content-Based Filtering" atau filtrasi berbasis konten. Metode ini berfokus pada analisis konten atau karakteristik lagu itu sendiri, seperti genre, tempo, instrumen, dan lirik, untuk merekomendasikan lagu-lagu yang sesuai dengan preferensi musik pengguna. Dalam sistem rekomendasi berbasis konten, kesesuaian antara karakteristik lagu dengan preferensi pengguna merupakan faktor kunci dalam menghasilkan rekomendasi yang relevan dan akurat[2].

Permasalahan perancangan ini adalah sejauh mana akurasi sistem rekomendasi dalam merekomendasikan lagu-lagu yang sesuai dengan selera musik pengguna. Dalam konteks teknologi informasi dan platform streaming musik yang semakin berkembang, sistem rekomendasi berbasis konten menjadi penting untuk membantu pengguna menemukan musik yang sesuai dengan preferensi mereka. Namun, tantangan yang dihadapi adalah bagaimana meningkatkan ketepatan rekomendasi agar pengguna dapat menerima saran musik yang lebih relevan dengan preferensi musik pribadi pengguna.

2. METODE PENELITIAN

Sistem yang dirancang ini menggunakan metode *content-based filtering* dan perhitungan kemiripan kosinus (*cosine similarity*) dan diimplementasikan dalam sebuah aplikasi web berbasis teknologi yang sesuai. Tujuan utama aplikasi ini adalah memungkinkan pengguna untuk melakukan pencarian, klasifikasi, dan mengetahui tingkat kesamaan antar item dalam konteks musik. Dengan demikian, sistem ini akan membantu pengguna dalam mengeksplorasi dan menikmati musik sesuai dengan preferensi mereka, sambil memberikan fitur tambahan untuk pengelolaan dan analisis yang diperlukan dalam konteks musik.



Gambar 1. Skema Perancangan

https://doi.org/10.24912/jsstk.v1i2.31037

Content-Based Filtering adalah metode dalam sistem rekomendasi yang memanfaatkan karakteristik atau konten dari item yang telah dikonsumsi oleh pengguna untuk merekomendasikan item yang serupa. Dalam konteks platform streaming musik seperti Spotify, metode ini menganalisis fitur-fitur musik dari lagu-lagu yang telah didengarkan oleh pengguna untuk menghasilkan rekomendasi lagu-lagu dengan karakteristik serupa. Tahapan utama dalam Content-Based Filtering melibatkan ekstraksi fitur, preprocessing teks, dan penggunaan teknik seperti *TF-IDF* (*Term Frequency-Inverse Document Frequency*) [3].

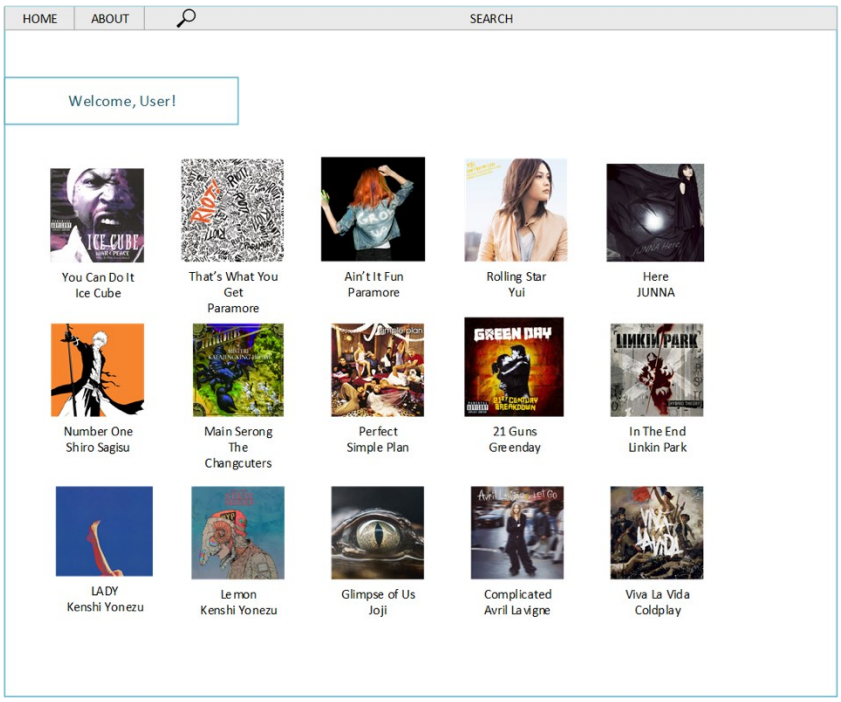
1. *Preprocessing*
   Preprocessing adalah langkah esensial dalam mengubah data yang tak terstruktur menjadi format yang lebih terstruktur, sehingga data tersebut dapat diolah lebih lanjut sesuai dengan kebutuhan. Dalam penelitian ini, data yang digunakan adalah lirik lagu. Tahap preprocessing mencakup pembersihan data untuk mencapai konsistensi dalam pola data sebelum melanjutkan proses selanjutnya. Ini dilakukan agar data lirik lagu dapat diolah dengan baik dan sesuai dengan kebutuhan analisis berikutnya. Berbagai tahapan preprocessing yang diterapkan mencakup *case folding, stemming,* dan *stopword removal.* Selanjutnya akan digunakan pembobotan *TF-IDF* lalu dihitung kemiripan nya menggunakan *Cosine similarity*.
2. *Case Folding*
   Proses dalam pemrosesan teks yang bertujuan untuk membuat semua huruf dalam teks menjadi huruf kecil (lowercase) atau, dalam beberapa kasus, huruf besar (uppercase). Tujuan utama case folding adalah untuk memastikan konsistensi dalam analisis teks, sehingga kata-kata yang sama dengan huruf besar dan huruf kecil tidak dianggap berbeda.
3. *Stemming*
   Tujuan dari stemming adalah untuk menghapus afiks (imbuhan) dari kata sehingga kata-kata yang memiliki akar kata yang sama akan dianggap setara, meskipun mereka memiliki imbuhan yang berbeda. Ini membantu dalam konsolidasi kata-kata yang memiliki makna yang sama dalam analisis teks.
4. *Stopword Removal*
   Proses menghapus kata-kata tertentu dari teks yang sering muncul namun tidak memiliki makna penting dalam analisis teks atau pemrosesan bahasa alami (NLP)[4]. Kata-kata ini disebut "stopwords" karena mereka "berhenti" atau diabaikan dalam analisis teks karena mereka tidak membawa informasi penting tentang konten teks yang sedang diproses.
5. *TF-IDF*
   Salah satu teknik yang digunakan dalam Content-Based Filtering untuk mengukur pentingnya suatu kata atau fitur dalam dokumen atau lagu. Teknik ini memberikan bobot lebih tinggi pada kata-kata yang muncul lebih sering dalam dokumen tertentu namun jarang muncul dalam dokumen lain[5]. Tujuan utama dari TF-IDF adalah untuk mengekstrak kata-kata yang penting atau khas dalam suatu dokumen dalam konteks korpus yang lebih besar. Dengan memanfaatkan TF-IDF, istilah-istilah atau kata-kata penting yang paling signifikan dalam suatu dokumen atau kumpulan teks bisa diidentifikasi dengan lebih baik. Rumus TF-IDF untuk menghitung bobot suatu kata *t* dalCosine Similarity dapat dibangun dengan langkah berikut. Pertama, perlu mengumpulkan data musik yang mencakup informasi seperti judul, genre, artis, tahun rilis, lirik, dan atribut audio. Data ini harus bersih dan bebas dari duplikat. Selanjutnya, ekstraksi fitur dari data musik, seperti penggunaan musik TF-IDF untuk lirik.

Selanjutnya, membuat profil pengguna dengan mengumpulkan preferensi musik mereka, seperti genre yang disukai, artis favorit, atau lagu-lagu tertentu. Dengan profil pengguna yang sudah ada, Anda dapat menghitung skor kesamaan cosine similarity antara profil pengguna dan setiap item dalam dataset. Cosine similarity adalah metrik yang akan membantu sistem memahami kesamaan antara preferensi pengguna dan item dalam dataset.

Setelah itu, mengurutkan item dalam dataset berdasarkan skor kesamaan cosine similarity secara menurun. Item dengan skor paling tinggi akan muncul pertama sebagai rekomendasi kepada pengguna. Anda dapat menampilkan beberapa rekomendasi teratas atau bahkan membuat daftar putar yang disesuaikan dengan preferensi pengguna.

Terakhir, mengimplementasikan sistem ini dalam bentuk aplikasi atau platform yang dapat diakses oleh pengguna. Diagram navigasi ini disusun untuk mengilustrasikan pengarahan antara modul-modul yang terdapat dalam sistem. Modul-modul yang terkandung dalam rancangan sistem ini meliputi modul beranda, modul search, modul detail aplikasi, dan modul admin.

1. Modul Beranda
   Modul Beranda merupakan laman awal dari situs web yang akan dikembangkan. Di dalam halaman beranda, pengguna akan menemukan daftar lagu yang dapat mereka pilih, serta kolom pencarian yang memungkinkan mereka mencari lagu sesuai dengan preferensi pribadi. Tampilan dapat dilihat pada Gambar 2.



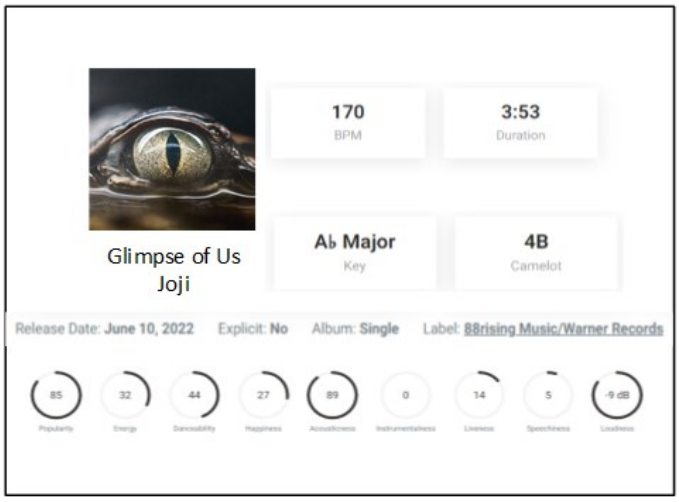
Gambar 2. Modul Beranda

2. Modul Search
   Modul search adalah halaman pencarian, di mana pengguna yang telah menggunakan kolom pencarian di halaman beranda akan dialihkan ke halaman ini. Di halaman pencarian ini, pengguna dapat melakukan pencarian melalui bidang yang disediakan dan memilih lagu berdasarkan hasil pencarian. Tampilan dapat dilihat pada Gambar 3.

Gambar 3. Modul Search

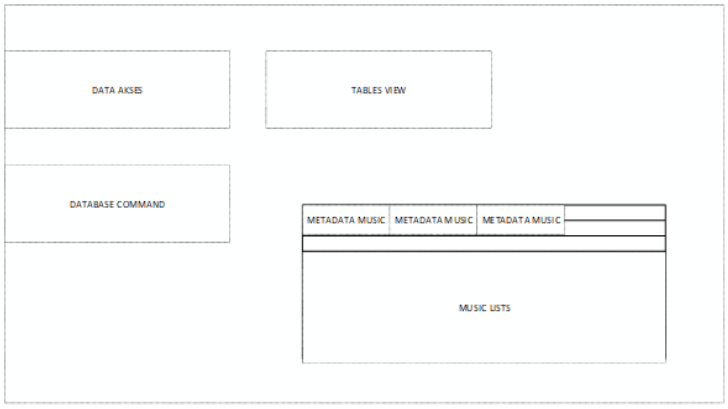
https://doi.org/10.24912/jsstk.v1i2.31037

3. Modul Detail Aplikasi
   Modul Aplikasi adalah detail lagu yang sudah dipilih dan akan menunjukkan detail dari lagu tersebut, seperti judul lagu, author, lama musik. Tampilan dapat dilihat pada Gambar 4.



Gambar 4. Modul Detail Aplikasi

4. Modul Admin
   Modul Admin adalah halaman admin yang hanya dapat diakses oleh pengguna dengan peran sebagai admin. Pada modul ini, pengguna dengan peran admin dapat melakukan perubahan data lagu dan melihat riwayat dari semua pengguna yang telah melakukan pencarian dan klasifikasi. Tampilan dapat dilihat pada Gambar 5.



Gambar 5. Modul Admin

Perancangan dilakukan menggunakan dataset lagu sebanyak 57.650 dan mengandung atribut artist, song, link, text. Pengujian metode memanfaatkan pemrograman Python dengan libraries Pandas dataframe dan Numpy. Dari hasil analisis terhadap metode Content-based filtering menunjukkan bahwa diperlukan atribut lagu yang detail untuk mendapatkan hasil rekomendasi yang baik.

Metode pengujian dilakukan mulai dari pembersihan kata-kata pada kolom text kemudian diberikan skor TF-IDF, setelah itu dilakukan perhitungan Cosine Similarity antara kata-kata yang diberikan pengguna dengan lirik lagu. Terakhir, akan diberikan rekomendasi yang sesuai. Hasil tampilan rekomendasi dapat dilihat pada Gambar 6 dan Gambar 7.(jangan ghambar dibawah, gambar 6)

Pada dataset sudah dilatih, dapat dilihat bahwa dimisalkan seorang pengguna ingin mencari lagu yang memiliki kemiripan dengan lagu "*Latest Flame",* dari program akan diberikan 10 rekomendasi lagu yang memiliki kesamaan yang mendekati dengan lagu "*Latest Flame".* Hasil rekomendasi sudah diurutkan dari yang memiliki kemiripan paling besar hingga paling kecil.

```
The 10 recommended songs for Latest Flame are:
Number 1:
Memphis Tennessee by Wanda Jackson with 0.242 similarity score
--------------------
Number 2:
Marching Into The Dark by John Legend with 0.229 similarity score
--------------------
Number 3:
Your Latest Trick by Dire Straits with 0.145 similarity score
--------------------
Number 4:
Media Overkill by Scorpions with 0.143 similarity score
--------------------
Number 5:
2 Days After Midnite by Wet Wet Wet with 0.137 similarity score
--------------------
Number 6:
Luck Of The Draw by Bonnie Raitt with 0.13 similarity score
--------------------
Number 7:
Flame by Horrible Histories with 0.126 similarity score
--------------------
Number 8:
Talked Myself Right Into It by Lynyrd Skynyrd with 0.114 similarity
score
--------------------
Number 9:
Love Me Like You Mean It by Bryan White with 0.111 similarity score
--------------------
Number 10:
The Girl From Yesterday by Eagles with 0.102 similarity score
--------------------
```

Gambar 6. Hasil Rekomendasi Program

https://doi.org/10.24912/jsstk.v1i2.31037

```
recommedations.recommend(recommendation)
```

```
The 10 recommended songs for Latest Flame are:
Number 1:
Memphis Tennessee by Wanda Jackson with 0.242 similarity score
--------------------
Number 2:
Marching Into The Dark by John Legend with 0.229 similarity score
--------------------
Number 3:
Your Latest Trick by Dire Straits with 0.145 similarity score
--------------------
Number 4:
Media Overkill by Scorpions with 0.143 similarity score
--------------------
Number 5:
2 Days After Midnite by Wet Wet Wet with 0.137 similarity score
--------------------
Number 6:
Luck Of The Draw by Bonnie Raitt with 0.13 similarity score
--------------------
Number 7:
Flame by Horrible Histories with 0.126 similarity score
--------------------
Number 8:
Talked Myself Right Into It by Lynyrd Skynyrd with 0.114 similarity score
--------------------
Number 9:
Love Me Like You Mean It by Bryan White with 0.111 similarity score
--------------------
Number 10:
The Girl From Yesterday by Eagles with 0.102 similarity score
--------------------
```

Gambar 7. Hasil Rekomendasi Pengguna 1

```
recommedations.recommend(recommendation2)
```

```
The 10 recommended songs for Home are:
Number 1:
Total Eclipse Of The Heart by One Direction with 0.332 similarity score
--------------------
Number 2:
Natural Woman by Kelly Clarkson with 0.323 similarity score
--------------------
Number 3:
That's How I Feel by Phil Collins with 0.321 similarity score
--------------------
Number 4:
60 Miles An Hour by New Order with 0.308 similarity score
--------------------
Number 5:
Road To Acceptance by Green Day with 0.306 similarity score
--------------------
Number 6:
I'll Be by Reba Mcentire with 0.306 similarity score
--------------------
Number 7:
She Makes Me (Feel Alright) by Journey with 0.293 similarity score
--------------------
Number 8:
18 by One Direction with 0.273 similarity score
--------------------
Number 9:
There You'll Be by Faith Hill with 0.272 similarity score
--------------------
Number 10:
Double Dare by Yo La Tengo with 0.259 similarity score
--------------------
```

Gambar 8. Hasil Rekomendasi Pengguna 2

Gambar 7 dan 8 menunjukan bahwa daftar rekomendasi antara pengguna akan berbeda, tergantung dengan kata kunci yang diberikan.

3. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan dari perancangan penerapan metode *Content-Based Filtering* dalam implementasi sistem rekomendasi musik, didapatkannya kesimpulan sebagai berikut:

1. Metode content-based filtering dapat diimplementasikan pada sistem rekomendasi musik.
2. Metode ini membantu pengguna menemukan lagu sesuai dengan preferensi musik pengguna dengan menganalisis karakteristik lagu, seperti genre, tempo, instrumen, dan lirik.
3. Implementasi menggunakan cosine similarity dapat memberikan rekomendasi yang personal dan akurat berdasarkan kemiripan karakteristik musik.

Saran

Untuk merancang aplikasi dengan tema serupa di masa mendatang dan meningkatkan kinerja dari metode yang telah diimplementasikan, berikut beberapa saran yang dapat diterapkan:

1. Tingkatkan personalisasi rekomendasi dengan mempertimbangkan preferensi dan riwayat mendengar pengguna secara lebih mendalam.
2. Menambahkan dataset lagu yang lebih banyak dari berbagai genre yang berbeda.

Ucapan Terima Kasih (*Acknowledgement*)

Puji dan syukur penulis panjatkan kehadirat Allah SWT, karena atas rahmat dan ridha-Nya penulis dapat menyelesaikan penyusunan makalah ini, dengan judul Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Musik.

Penulis ingin menyampaikan banyak terima kasih kepada berbagain pihak, kepada Viny Christanti Mawardi, S.Kom., M.Kom., TRI SUTRISNO S.Si., M.Sc., dan rekan-rekan angkatan. Akhir kata, penulis sampaikan dan ucapkan terimakasih yang sebesar-besarnya kepada semua pihak yang telah membantu dan mendukung dalam penyusunan skripsi ini.

REFERENSI

Christanti M., V., Kurniawan, F. and Tony. (2011). Automatic music classification for Dangdut and campursari using Naïve Bayes, *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics,* Bandung, Indonesia, pp. 1-6, doi: 10.1109/ICEEI.2011.6021738.

Daniel, Mulyawan B., Sutrisno, T. (2022). Pembuatan Aplikasi E-Commerce Berbasis Web Dengan Fitur Rekomendasi Menggunakan Metode Content-Based Filtering, *Jurnal Ilmu Komputer dan Sistem Informasi*, 10(1).

Data Science Tutorials. (n.d.). *TF-IDF and Cosine Similarity in Machine Learning*. Retrieved September 6, 2023, from https://dotnettutorials.net/lesson/tf-idf-and-cosinesimilarity-in-machine-learning/?expand_article=1

Putra, A.I., Santika, R.R. (2020). Implementasi Machine Learning dalam Penentuan Rekomendasi Musik dengan Metode Content-Based Filtering. *Edumatic: Jurnal Pendidikan Informatika*, 4(1), 121–130. https://doi.org/10.29408/edumatic.v4i1.2162

Ula, N. (2021). *Sistem Rekomendasi Lagu Dengan Metode Content Based Filtering Berbasis Website.* [Bachelor thesis, Telkom University]. Telkom University Repository

https://doi.org/10.24912/jsstk.v1i2.31037