

SISTEM REKOMENDASI LAGU DENGAN METODE CONTENT-BASED FILTERING BERBASIS WEBSITE

WEB-BASED SONG RECOMMENDATION SYSTEM USING CONTENT-BASED FILTERING

Naziratul Ula¹, Casie Setianingsih², Ratna Astuti Nugrahaeni³

^{1,2,3} Universitas Telkom, Bandung

naziraaatulula@student.telkomuniversity.ac.id¹, setiacasie@telkomuniversity.ac.id²,

ratnaan@telkomuniversity.ac.id³

Abstrak

Lagu adalah salah satu media hiburan yang tidak pernah luput oleh zaman. Bagi penikmatnya, lagu adalah suatu hal yang sangat penting karena lagu merupakan suatu hiburan yang akan didengarkan sesuai dengan suasana hati mereka. Dikarenakan sangat banyak lagu yang tersedia, membuat banyak orang sulit dalam memilih lagu yang ingin didengar. Salah satu yang membuat banyak orang bingung adalah banyaknya lagu dengan jenis *genre* yang berbeda – beda. Suasana hati juga mempengaruhi jenis lagu yang akan didengarkan. Sistem rekomendasi adalah sebuah sistem yang mengacu pada memprediksi sejumlah *item* atau data untuk pengguna di masa mendatang, kemudian dijadikan rekomendasi *item* paling teratas. Sistem rekomendasi memberikan informasi yang relevan dengan apa yang dicari oleh pengguna. Pada penelitian ini, penulis menggunakan metode *Content-based filtering* untuk mencari rekomendasi lagu. Konten yang digunakan adalah lirik lagu. Algoritma TF-IDF digunakan untuk mencari nilai bobot *term/kata* pada tiap dokumen dan kemudian nilai tersebut digunakan sebagai variabel pada *Cosine similarity* untuk mencari kesamaan antar dokumen. Keluaran dari tugas akhir ini adalah sistem dapat memberi keluaran berupa rekomendasi 10 lagu teratas berbahasa Indonesia. Hasil pengujian dari tugas akhir ini adalah sistem sudah berjalan dengan baik dengan hasil uji alpha sebesar 100%.

Kata kunci : *Content-based filtering*, TF-IDF, *Cosine similarity*

Abstract

Songs are one of the entertainment media that never escape from the times. For the audience, the song is a very important thing because the song is an entertainment that will be listened to according to their mood. Because there are so many songs available, it makes it difficult for many people to choose the song they want to listen to. One thing that makes many people confused is that there are many songs with different types of genres. Mood also affects the type of song you will listen to. A recommendation system is a system that refers to predicting a number of items or data for users in the future, then being recommended as the top item. The recommendation system provides information that is relevant to what the user is looking for. In this study, the author uses the Content-based filtering method to find song recommendations. The content used is song lyrics. The TF-IDF algorithm is used to find the value of term/word weight in each document and then that value is used as a variable in Cosine similarity to find similarities between documents. The output of this final project is the system can provide output in the form of recommendations for the top 10 songs in Indonesian. The result of this final project is the system has been running well with alpha test result of 100%.

Keywords : *Content-based filtering*, TF-IDF, *Cosine similarity*

1. Pendahuluan

Lirik lagu juga merupakan salah satu hal yang sangat penting selain musik yang mengiringinya. *Mood* pada lagu juga dapat dilihat dari lirik. Lagu dengan suasana sedih misalnya, lirik yang digunakan juga akan menggunakan kata – kata yang mewakili perasaan sedih, begitu juga lainnya. Penulis menggunakan lirik sebagai konten dalam penelitian ini agar pengguna dapat mencari lagu dengan jenis lirik yang serupa.

Melihat permasalahan ini, dibutuhkan sebuah perangkat lunak yang dapat membantu *user* dalam mencari rekomendasi lagu. Sistem rekomendasi adalah sebuah sistem yang mengacu pada memprediksi sejumlah *item* atau data untuk pengguna di masa mendatang, kemudian dijadikan rekomendasi *item* paling teratas [1]. Sistem rekomendasi memberikan informasi yang relevan dengan apa yang dicari oleh pengguna. Sistem rekomendasi dirasa mampu membantu penelitian ini dalam memberikan rekomendasi lagu kepada masyarakat. Dalam sistem rekomendasi, ada beberapa metode yang dapat digunakan diantaranya *Collaborative filtering*, *Content-based filtering* dan gabungan keduanya yaitu *Hybrid*.

Penulis menggunakan metode *Content-based filtering* untuk membuat sistem rekomendasi lagu. Konten yang akan digunakan dalam penelitian ini adalah lirik lagu. Algoritma yang digunakan untuk menghitung kesamaan pada setiap lagu adalah *Cosine similarity*. Dokumen yang digunakan merupakan dokumen lirik lagu. Untuk mendapat hasil rekomendasi, ada beberapa tahap yang harus dilakukan. Tahap awal adalah proses pembersihan dokumen lirik lagu dengan proses *preprocessing*. Ada beberapa tahap yang dilalui pada proses *preprocessing* untuk mendapatkan data yang lebih bersih, yaitu *case folding*, *stemming* dan *stopword removal*. Setelah proses tersebut, maka harus dilakukan pembobotan dengan TF-IDF yang lalu dihitung jarak kemiripannya dengan *Cosine similarity*. *Similarity* akan dilihat dari tiap lirik lagu,

sehingga nilai *similarity* ini yang akan digunakan untuk mencari rekomendasinya sehingga sistem dapat mengeluarkan *output* berupa rekomendasi lagu kepada pengguna.

2. Landasan Teori

2.1. Lagu

Lagu merupakan gabungan antara musik yang terdiri dari nada – nada dan lirik yang memiliki makna. Lirik lagu merupakan ekspresi seseorang tentang suatu hal yang sudah dilihat, didengar maupun dialaminya. Dalam mengekspresikan pengalamannya, penyair atau pencipta lagu melakukan permainan kata-kata dan bahasa untuk menciptakan daya tarik dan kekhasan terhadap lirik atau syairnya [2].

Seiring berkembangnya zaman, lagu yang tersedia menjadi semakin banyak. Terdapat berbagai macam *genre* dan jenis lagu. Lagu yang terbentuk dari hubungan antara unsur musik dengan unsur syair atau lirik lagu merupakan salah satu bentuk komunikasi massa [2]. Pencipta lagu pasti menyelipkan pesan – pesan pada lagu yang diciptakannya. Pesan didalamnya pun dapat memiliki berbagai macam arti. Pesan – pesan tersebut dituangkan kedalam lirik maupun nada yang digunakan.

2.2. Preprocessing

2.2.1. Case Folding

Case folding digunakan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil. Sehingga hanya huruf ‘a’ sampai ‘z’ yang diambil, karakter selain huruf akan dihilangkan. Begitu juga dengan huruf – huruf kapital.

2.2.2. Stemming

Stemming merupakan proses menghilangkan imbuhan pada kata, contohnya ber, me, an. Sehingga kata – kata pada dokumen menjadi bentuk kata dasar.

2.2.3. Stopword Removal

Filtering atau biasa disebut *stop-word removal* adalah proses pada *preprocessing* yang berguna untuk menghilangkan kata sambung, kata penghubung atau kata umum lainnya [3]. Sehingga yang tersisa hanya kata – kata penting dan kata – kata yang tidak penting akan dibuang.

2.3. Sistem Rekomendasi

Sistem rekomendasi secara khusus merupakan suatu sistem pemberian saran kepada penggunanya dan bersifat personal, berbeda untuk semua pengguna sistem. Bertujuan untuk memberikan informasi yang sekiranya menarik bagi pengguna dan membantu pengguna memutuskan apa yang akan disukai olehnya. Perkiraan informasi ini bersifat personal yang didasarkan atas profil dari pengguna sistem. Profil pengguna umumnya didasarkan atas penilaian menarik-tidaknya suatu informasi yang pernah dibaca oleh pengguna [4].

Sistem rekomendasi memiliki beberapa metode yang umum digunakan yaitu, *content-based filtering*, *collaborative filtering* (*User-based filtering* dan *Item-based filtering*) dan *hybrid* yang merupakan penggabungan dari kedua metode tersebut. Pada *collaborative filtering*, metode dibagi lagi menjadi dua yaitu *user-based* yang mana pengguna memainkan peran penting dan *item-based* yang mana *item* yang lebih memainkan peran penting [5].

2.4. Content-based Filtering

Content-based filtering merupakan metode yang memanfaatkan informasi dari beberapa *item* atau data yang kemudian direkomendasikan kepada pengguna sebagai referensi yang terkait dengan informasi yang digunakan sebelumnya. Pada kasus *Content-based filtering*, rekomendasi yang diberikan berdasarkan informasi tentang konten *item* tertentu [6]. Contohnya, jika *user* menyukai A dan B mirip dengan A, maka *user* mungkin akan menyukai B [7]. Metode *Content-based filtering* membentuk profil penggunanya berdasarkan atribut pembentuk suatu *item* [8].

Algoritma pada metode *Content-based filtering* dapat dijelaskan pada tahap-tahap berikut ini:

1. Suatu *item* barang dibagi-bagi berdasarkan vektor komponen pembentuknya.
2. Pengguna akan memberikan penilaian berupa suka atau tidak suka pada item.
3. Sistem akan membentuk profil pengguna berdasarkan bobot vektor komponen suatu *item*. Pembuatan profil pengguna dapat menggunakan algoritma TF-IDF (*term frequency-invers document frequency*). TF merupakan jumlah *term* dalam sebuah dokumen. Sedangkan IDF dapat dihitung dengan rumus :

$$idf_i = \log\left(\frac{n}{df_i}\right) \quad (2.1)$$

$$W_{td} = tf \times IDF \quad (2.2)$$

$$W_{td} = tf \times \log \frac{N}{df} \quad (2.3)$$

Dimana:

d : dokumen ke-d

t : kata ke-t dari kata kunci

W : bobot dokumen ke-d terhadap kata ke-t

tf : banyak kata yang dicari pada sebuah dokumen

IDF : Inversed Document Frequency

N : jumlah total dokumen

df : banyak dokumen yang mengandung *term/kata* t

2.4.1. Cosine Similarity

Cosine similarity adalah salah satu metode yang paling populer yang sering diterapkan pada teks dokumen untuk keperluan temu balik informasi dan klastering [9]. *Cosine similarity* menghitung vektor yang terkait antara satu dengan yang lain [10]. Nilai *similarity* yang dihasilkan adalah antara nilai 0 dan nilai 1 [11]. Nilai 0 menandakan *user profile* dan *item* sangat tidak relevan sedangkan nilai 1 adalah sebaliknya. Untuk mencari *cosine similarity* antar dokumen dapat dinyatakan pada persamaan berikut:

$$SIM(\vec{t}_a, \vec{t}_b) = \frac{\vec{t}_a \cdot \vec{t}_b}{|\vec{t}_a| \times |\vec{t}_b|} \quad (2.2)$$

Dimana :

ta = vektor a

tb = vektor b

Dimana ta^{\rightarrow} dan tb^{\rightarrow} adalah vektor multi dimensional dari kumpulan istilah $T = \{t_1, t_2, t_3, \dots, t_n\}$ [9].

3. Perancangan

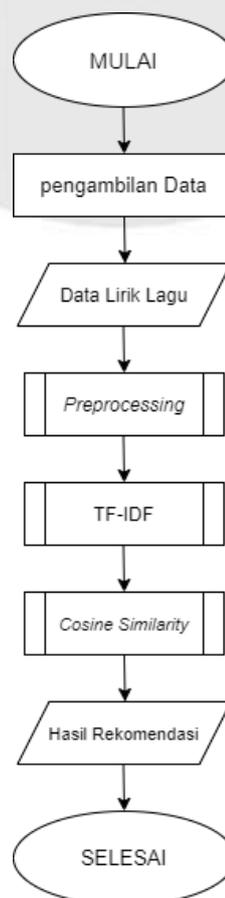
3.1. Gambaran Umum Sistem

Sistem yang akan dibangun adalah sistem yang mampu memberi rekomendasi lagu berbahasa Indonesia dengan mengimplementasikan salah satu metode sistem rekomendasi. Metode yang digunakan adalah *Content-based filtering* dengan algoritma *Cosine similarity*.

Data yang digunakan merupakan data lirik lagu berbahasa Indonesia yang diambil dari beberapa situs di internet. Data lirik lagu ini akan melalui tahapan *preprocessing*, lalu selanjutnya data yang telah di *preprocessing* akan dilakukan pembobotan dan terakhir data tersebut dihitung nilai *similarity* menggunakan algoritma *Cosine similarity*.

3.2. Perancangan Sistem Rekomendasi

Dibawah ini merupakan perancangan sistem yang akan dibuat.

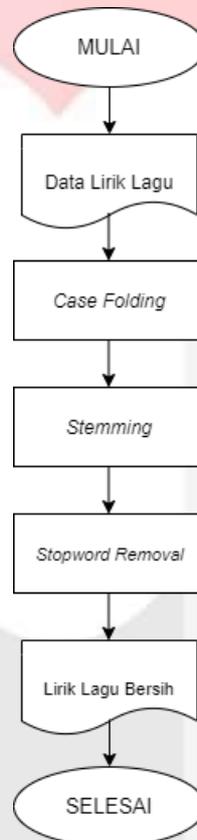


Gambar 3.1 Flowchart Perancangan Sistem

Dimulai dari pengambilan data penyanyi, judul dan lirik lagu berbahasa Indonesia. Data diambil secara manual melalui situs website. Data lirik lagu akan digunakan untuk proses *preprocessing* untuk mendapatkan data yang lebih bersih. Selanjutnya data akan melalui tahap pembobotan TF-IDF, lalu tahap terakhir adalah mencari nilai *similarity* menggunakan algoritma *Cosine similarity*.

3.2.1. *Preprocessing*

Preprocessing adalah tahap pembersihan data agar memiliki pola yang sama sebelum dilakukan proses selanjutnya. Data yang digunakan pada tugas akhir ini adalah data lirik lagu. Tahap *preprocessing* yang digunakan antara lain :



Gambar 3.3 Flowchart preprocessing

Pada **Gambar 3.3** menggambarkan alur dari proses *preprocessing* dimulai dari data lirik lagu mentah yang kemudian melalui tahap *case folding*, *stemming* dan *stopword*. Berikut penjelasan lebih detail dari tahap - tahap pada *preprocessing* :

3.2.2. *Case Folding*

Case folding merupakan tahap pengubahan semua huruf kapital menjadi huruf kecil. Masukan pada proses *case folding* merupakan data mentah lirik lagu yang telah dikumpulkan sebelumnya.

3.2.3. *Stemming*

Stemming merupakan suatu proses mentransformasi kata-kata yang terdapat dalam suatu dokumen menjadi akar kata atau dasar kata dengan aturan-aturan tertentu, atau dapat dianggap sebagai pembuangan imbuhan pada kata. *Stemming* memiliki aturan tersendiri untuk Bahasa yang digunakan. Untuk proses bahasa Indonesia memiliki struktur imbuhan yang kompleks sehingga lebih susah untuk diolah. Aturan *stemming* yang digunakan pada perancangan sistem kali ini menggunakan *tools* yang sudah disediakan dari Sastrawi.

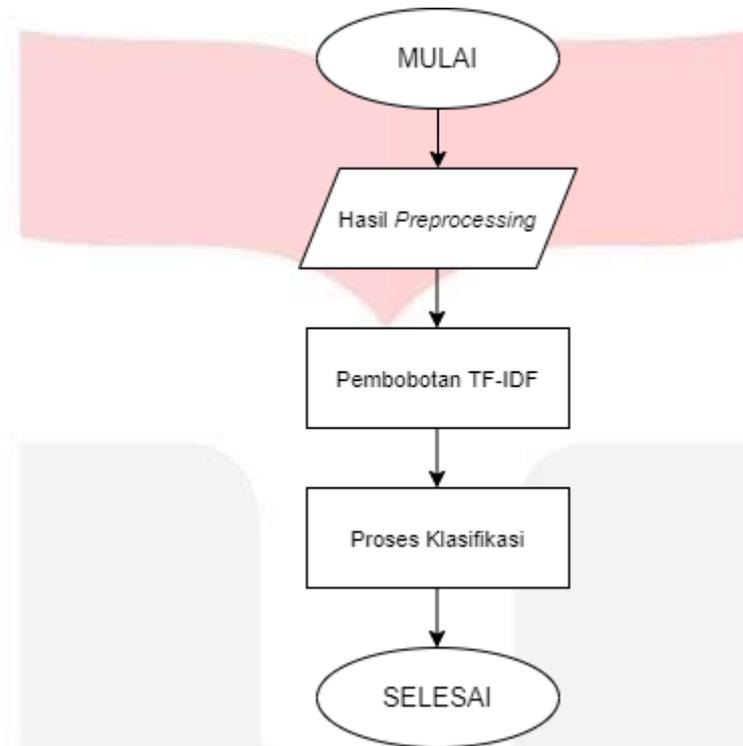
3.2.4. *Stopword Removal*

Stopword removal merupakan proses menghilangkan kata pada dokumen yang dianggap tidak memiliki arti atau tidak mengandung emosi. Dokumen yang digunakan pada tugas akhir ini merupakan dokumen lirik lagu. Pembuangan kata dilakukan dengan mencocokkan kata pada kamus *stopword removal*.

Proses *stopword removal* juga menggunakan kamus *stopword removal* yang berisi kata – kata yang dianggap tidak mengandung emosi dan tidak mencirikan kata emosi.

3.2.5. **TF-IDF**

TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan untuk menghitung pembobotan kata dalam dokumen. Perhitungan TF-IDF bertujuan untuk memberikan bobot nilai pada sebuah kata yang dilihat dari seberapa sering kata tersebut muncul didalam suatu dokumen.



Gambar 3.4 Flowchart TF-IDF

3.2.6. Cosine Similarity

Algoritma rekomendasi yang digunakan pada tugas akhir ini adalah *Cosine similarity*. *Cosine similarity* digunakan bertujuan untuk menghitung kemiripan vektor – vektor pada dokumen. Selanjutnya dokumen dengan nilai *similarity* terdekat akan dijadikan rekomendasi.



Gambar 3.5 Flowchart Cosine Similarity

Pada Gambar 3.4 merupakan flowchart dari algoritma *Cosine similarity*. Masukan untuk proses perhitungan *Cosine similarity* adalah nilai yang didapat dari perhitungan bobot *term* pada TF-IDF untuk dijadikan bentuk vektor.

4. Implementasi dan Pengujian

4.1. Pengumpulan Dataset

Data yang dibutuhkan untuk tugas akhir ini adalah data lirik lagu, beserta artis, dan judul dan url youtube. Sedangkan jenis *genre* dan *mood* hanya sebagai keterangan dari *output* rekomendasinya agar pengguna tahu lagu

yang direkomendasikan untuknya memiliki jenis *genre* dan *mood* apa. Data artis, judul, *genre* dan *mood* diambil secara manual melalui *website* <https://www.chosic.com/> . sedangkan untuk lirik lagu diambil secara manual melalui beberapa *website* yaitu <https://www.kapanlagi.com/>, <https://www.lyricfind.com/>, dan <https://www.musixmatch.com/>

4.2. Preprocessing Dataset

Preprocessing dataset meliputi mengubah semua huruf kapital menjadi huruf kecil dengan case folding, mentransformasi kata-kata yang terdapat dalam suatu dokumen menjadi akar kata atau dasar kata dengan aturan-aturan tertentu. menghilangkan kata pada dokumen yang dianggap tidak memiliki arti atau tidak mengandung emosi.

4.3. Pengujian Sistem

Pengujian yang dilakukan pada penelitian ini menggunakan pengujian *alpha* dan *beta* untuk sistem yang telah dibuat. Pengujian *alpha* adalah dengan cara melihat performansi pada sistem dengan *black box testing* dan pengujian *beta* dengan menyebarkan kuisioner secara langsung kepada pengguna. Setelah melakukan pengujian *alpha* dan *beta*, sistem telah dinyatakan berhasil dan dapat berjalan dengan baik.

5. Kesimpulan dan Saran

5.1. Kesimpulan

Dari pengujian yang telah dilakukan dapat disimpulkan bahwa sistem yang dibuat telah berhasil menjawab tujuan dari penelitian dan dapat ditarik kesimpulan sebagai berikut :

1. Dari hasil pengujian alpha memberikan hasil bahwa sistem dan fitur pada *website* rekomendasi lagu dapat berjalan dengan baik dan berjalan sesuai dengan input pengguna.
2. Sistem yang dirancang berhasil memberikan rekomendasi lagu berdasarkan nilai bobot *term* dan *similarity* dari lirik lagu.
3. Sistem sudah dapat berjalan dengan baik dengan hasil uji alpha sebesar 100%.

5.2. Saran

Adapun saran dari penulis untuk penelitian dengan tema yang sama seperti ini dimasa mendatang, serta diharapkan dapat meningkatkan kinerja dari penggunaannya sebagai berikut :

1. Dapat membuat sistem rekomendasi lagu dengan konten yang berbeda dan lebih akurat.
2. Menambahkan *dataset* yang lebih banyak dan lagu yang digunakan tidak hanya lagu berbahasa Indonesia.

Referensi

- [1] "Data Folks Indonesia," medium.com, 11 July 2019. [Online]. Available: <https://medium.com/data-folks-indonesia/recommendation-system-dengan-python-definisi-part-1-71154dc3f700>. [Accessed 27 November 2020].
- [2] R. Hidayat, "Analisis Semiotika Makna Motivasi Pada Lirik Lagu "Laskar Pelangi" Karya Nidji," *Jurnal Ilmu Komunikasi*, vol. 2, no. 1, pp. 243-258, 2014.
- [3] R. Y. H. R. R. M. P. Randy Cahya Wihandika, "Implementasi Metode K-Nearest Neighbour Dengan Pembobotan TF.IDF.ICF Untuk Kategorisasi Ide Kreatif Pada Perusahaan," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 4, no. 2, pp. 97-103, 2017.
- [4] P. S. Adi, "SISTEM REKOMENDASI NILAI MATA KULIAH MENGGUNAKAN METODE CONTENT-BASED FILTERING," Yogyakarta.
- [5] S. P. H. G. M.H. Yap, "A Novel Hybrid based Recommendation System based on Clustering and Association Mining," in *Tenth International Conference on Sensing Technology*, UK, 2016.
- [6] T. R. J. R. Piotr Woldan, "A Content-Based Recommendation System Using Neuro-Fuzzy Approach," in *International Conference on Fuzzy System*, Rio de Janeiro, Brazil, 2018.
- [7] F. F. L. P. Tiffany Zhan, "T-RECSYS: A Novel Music Recommendation System Using Deep Learning," in *International Conference On Consumer Electronics*, USA, 2019.

- [8] T. B. R. F. Iwan Syarif, "Sistem rekomendasi content based filtering menggunakan algoritma apriori," in *Konferensi Nasional Sistem Informasi, STMIK Atma Luhur Pangkalpinang*, 2018.
- [9] C. S. D. Prasetya, "Sistem rekomendasi pada e-commerce menggunakan k-nearest neighbor," *Jurnal Teknologi Informasi dan Ilmu Komputer*, pp. 194-200, 2017.
- [10] A. A. Minakshi Chauchan, "Similarity measures used in recommender system: A study," *International Journal of Engineering Technology Science and Research*, vol. 4, no. 6, pp. 619-626, 2017.
- [11] G. R. E. R. S. M. Ade Romadhony, ST., MT., "Analisis dan implementasi House recommendation menggunakan content-based recommender system".