OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 10, Oktober 2024
ISSN 2828-2442 (media online)
Hal 2503-2509

Sistem Rekomendasi Drama dan Film Berbasis *Website* Dengan Metode *Content Based Filtering*

Yusniar[1], Yolen Perdana Sari[1*]

[1]Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia
Email: [1]yusniarjung@gmail.com, [2*]dosen01705@unpam.ac.id
(* : coressponding author)

Abstrak−Perkembangan dalam indstri sineas sangat pesat dari masa ke masa. Ribuan judul film dan drama muncul dalam setahun dari berbagai genre, pemain dan plotnya. Sistem rekomendasi akan membantu pengguna dalam memutuskan pilihan berdasarkan data yang akan di proses. Dalam penelitian ini, system rekomendasi drama dan film akan dibangun dengan metode content based filtering dengan perhitungan TF-IDF serta *cosine similarity* melalui data seperti judul, genre, sinopsis serta pemain sehingga pengguna dapat dengan mudah mencari tontonan berdasarkan jenis drama atau film yang disukai.

Kata Kunci: Sistem Rekomendasi, Film, Drama, Content Based Filtering

Abstract*−Developments in the film industry have been very rapid from time to time. Thousands of film and drama titles appear in a year from various genres, players and plots. The recommendation system will assist users in making choices based on the data to be processed. In this study, a drama and film recommendation system will be built using the content-based filtering method with TF-IDF calculations and cosine similarity through data such as title, genre, synopsis and cast so that users can easily search for shows based on the type of drama or film they like.*

Keywords*: Recommendation System, Film, Drama, Content Based Filtering*

1. PENDAHULUAN

Saat ini, perkembangan internet di dunia sangat pesat, Internet telah menjadi sumber kebutuhan pokok yang digunakan sehari-hari, mulai dari sektor Pendidikan, industri, maupun hiburan. Salah satu hiburan yang sampai saat ini masih banyak di nikmati oleh masyarakat adalah film maupun drama. Industri perfilman saat ini dapat di nikmati lewat layar lebar maupun aplikasi penyedia layanan streaming online yang menyajikan ratusan film maupun drama. Jumlah film yang di rilis di Indonesia pada tahun 2021 mencapai 106 judul film dengan berbagai genre dan tayang di layar lebar maupun platform streaming video on-demand streaming video seperti Netflix, Disney+ Hotstar, WeTv dan Viu.

Banyaknya film dan drama yang di rilis pun semakin menambah informasi tentang sinema tersebut di internet, mulai dari genre, pemain, sinopsis maupun negara asal sinema tersebut. Hal ini lah yang sekarang mulai jadi permasalahan bagi seseorang yang ingin mencari hiburan dan ingin memilih film ataupun drama. Sebagian orang sudah tahu akan menonton apa berdasarkan ketertarikannya. Tetapi, bagi orang yang mencari hiburan dalam waktu senggangnya, dengan begitu banyaknya pilihan dan jenis sinema yang ditawarkan menjadi lebih sulit, menjadikan mereka terpaksa memilih secara acak apa yang akan di tonton. Atau bahkan memilih tidak menontonnya sama sekali.

Content-based filtering dapat digunakan untuk memberikan rekomendasi item yang mirip dengan yang diminati pengguna di waktu sebelumnya, dengan cara suatu item dihitung berdasarkan atribut yang berkaitan dengan item yang dibandingkan. Perhitungan kemiripan yang dapat digunakan untuk menunjukan rekomendasi yang tepat yaitu dengan perhitungan TF-IDF serta *cosine similarity*. Dengan menggunakan dataset drama dan film yang berasal dari situs penyedia data, Kaggle. Penelitian ini akan berfokus pada *Content-based Filtering* dalam pembuatan website rekomendasi.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 10, Oktober 2024
ISSN 2828-2442 (media online)
Hal 2503-2509

2. METODOLOGI PENELITIAN

2.1 Sistem Rekomendasi

Sistem rekomendasi merupakan alat personalisasi yang mencoba memberikan pelayanan bagi penggunanya berupa daftar informasi sesuai dengan selera dan keinginan pengguna. Sistem rekomendasi akan menyimpulkan kesukaan pengguna terhadap suatu hal dan menganalisisnya melalui data pengguna, informasi pengguna lainnya dan informasi tentang lingkungannya (Sebastia dkk, 2009).

Sistem Rekomendasi adalah sebuah system yang akan merekomendasikan suatu item kepada penggunanya berdasarkan preferensi dari pengguna tersebut. Sistem akan mengelola data pengguna dan merekomendasikan item yang paling sesuai sesuai informasi yang ada dalam sistem.

2.2 *Content Based Filtering*

Content-Based Filtering merupakan hasil penelitian penyaringan dari sebuah informasi dalam sistem berbasis konten. Sistem rekomendasi dibuat untuk memahami kebutuhan dan preferensi pengguna. Nantinya informasi digabungkan dengan log dari interaksi pengguna sebelumnya. Lalu, sistem rekomendasi akan mencocokan informasi pengguna dengan informasi suatu produk yang ada di dalam database(Budi Utomo dan Yohanes Suhari, 2013). *Content-based Filtering* pada umumnya digunakan untuk merekomendasikan item yang berbasis teks. Dalam rekomendasi drama dan film nantinya berupa kata kunci seperti genre, tag, sinopsis, maupun pemain.

2.3 Pengumpulan Data

Pada penelitian ini, data yang digunakan berasal dari situs penyedia data, Kaggle. Dataset yang digunakan dibagi menjadi dua, yaitu data film dan data drama. Dataset yang diunduh berformat .csv(*comma separated values*) yang berisi 5000 data film serta 1500 data drama. Data-data tersebut berdasarkan list film dan drama yang berada pada situs TMDB. Data film serta drama tersebut memiliki atribut: *tmdbid, title, genre, overview, vote, cast, budget dan popularity*.

Kedua dataset ini akan melalui beberapa tahapan, yaitu data *preprocessing* dimana data akan dibersihkan guna menciptakan data yang lebih siap dipakai, kemudian data akan dihitung bobot stiap atributnya menggunakan algoritma TF-IDF dan selanjutnya akan dihitung menggunakan *cosine similarity* untuk membuat rekomendasi film dan drama.

Pada tahap *preprocessing* data, data akan dipilah kembali atributnya, dengan membuang atribut yang tidak diperlukan dalam membuat pengolahan data. Dalam penelitian ini, tahap preprocessing data akan terbagi menjadi 3 tahapan, yaitu *case folding, tokenizing, serta filtering(stopword removal).*

Gambar 1. Tahap *Preprocessing* Data

Perhitungan TF-IDF (*Term Frequency-Inverse Document Frequency*) biasanya digunakan sebagai metode untuk melakukan pengukuran pada deskripsi suatu item yang bersifat tekstual. Dengan menggunakan perhitungan ini, maka kita dapat mengukur bobot dari term yang muncul dalam atribut. persamaan untuk menghitung bobot TF-IDF dapat dilihat dalam persamaan(1).

$$W_{i,j} = tf_{i,j}\ log\ \frac{N}{dfi} \qquad (1)$$

Lalu perhitungan selanjutnya menggunakan cosine similarity, yaitu metode untuk mengukur kemiripan antara dua teks atau bobot yang digunakan. Cosine similarity bertujuan untuk menghitung setiap sudut dari dua atrbitu dalam vector shape dan menghitung persamaannya, dapat dilihat pada persamaan (2).

$$Sim(q, \mathbf{d}_j) = \frac{q\ x\ \mathbf{d}_j}{|q|\ x\ |\mathbf{d}_j|} = \frac{\sum_{i=1}^{n} W_{i,q}\ x\ W_{i,j}}{\sqrt{\sum_{i=1}^{t}(W_{i,q})^2}\sqrt{\sum_{i=1}^{t}(W_{i,j})^2}} \qquad (2)$$

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 10, Oktober 2024
ISSN 2828-2442 (media online)
Hal 2503-2509

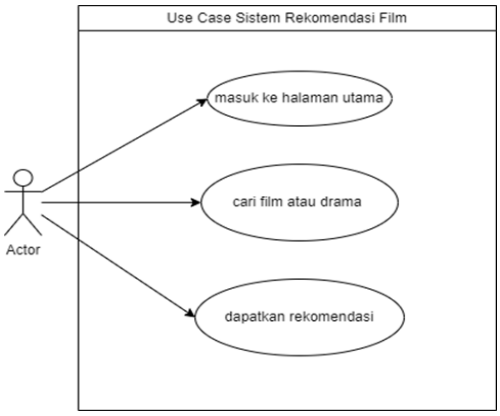
3. ANALISA DAN PEMBAHASAN

3.1 Analisa Sistem

Pada tahap ini, peneliti akan menentukan apa yang diperlukan dalam membuat system serta bagaimana sistem akan dibuat dan membuat proses kerja dari sistem. Selain itu Analisa memungkinkan peneliti untuk mengembangkan sistem dan menentukan cara dalam menyelesaikan masalah yang ada dan memprediksi hambatan yang kemungkinan muncul di kemudian hari untuk perbaikan kedepannya.. Dalam penelitian ini, peneliti menggunakan *framework* Streamlit dan Bahasa pemrograman pyhon untuk membuat kode program.

3.2 *Use Case Diagram*

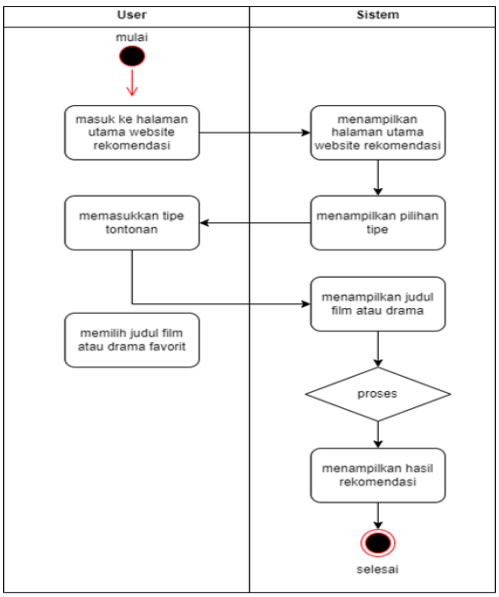
Proses yang ada dalam system akan digambarkan dalam bentuk *use case diagram*, dimana dapat melihat gambaran dari keseluruhan hubungan antara aktor dengan sistem. *Use Case* yang ada pada sistem akan ditunjukan pada gambar 2.



Gambar 2. *Use Case Diagram*

3.3 *Activity Diagram*

Dalam activity diagram, akan ditunjukan bagaimana hubungan antara aktor dengan sistem rekomendasi film akan berjalan dan menghasilkan suatu rekomendasi. Gambaran activity diagram akan ditunjukan pada gambar 3.



Gambar 3. *Activity Diagram*

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 10, Oktober 2024
ISSN 2828-2442 (media online)
Hal 2503-2509

3.3 Perhitungan Kemiripan

Dalam penelitian ini akan digunakan perhitungan kemiripan dengan algoritma TF-IDF untuk memberikan bobot pada setiap term, dalam hal ini adalah penggalan kata pada overview drama. Lalu perhitungan kedua akan menggunakan cosine similarity. Berikut ini contoh kalimat dalam overview yang sudah melalui tahap preprocessing akan diberikan bobot dengan TF-IDF.

a. Kalimat 1 : " love and friendship between high school students"
b. Kalimat 2 : "high school student had complex appearance."
c. Kalimat 3 : "high school student accused committing violence."

Perhitungan TF-IDF dari kalimat overview diatas dapat dilihat pada Tabel 1.

Tabel 1. Perhitungan TF-IDF

Kata	Doc 1	Doc 2	Doc 3	df	idf	Tf.idf Doc 1	Tf.idf Doc 2	Tf.idf Doc 3
Love	1	0	0	1	Log(3/1)=0.4771	0.4771	0	0
high	1	1	1	3	Log(3/3)=0	0	0	0
School	1	1	1	3	Log(3/3)=0	0	0	0
had	0	1	0	1	Log(3/1)=0.4771	0	0.4771	0
complex	0	1	0	1	Log(3/1)=0.4771	0	0.4771	0
appearance	0	1	0	1	Log(3/1)=0.4771	0	0.4771	0
student	1	1	0	2	Log(3/2)=0.1760	0.1760	0.1760	0
friendship	1	0	0	1	Log(3/1)=0.4771	0	0.4771	0
between	1	0	0	1	Log(3/1)=0.4771	0.4771	0	0
accused	0	0	1	1	Log(3/1)=0.4771	0	0	0.4771
comutting	0	0	1	1	Log(3/1)=0.4771	0	0	0.4771
violance	0	0	1	1	Log(3/1)=0.4771	0	0	0.4771

Selanjutnya perhitungan cosine similarity berguna untuk mengukur tingkat kemiripan antara dua vector. Cosine similarity nantinya akan menentukan kemiripan antara satu dara drama atau film dengan data drama dan film lainnya. Contoh perhitungan cosine similarity dilihat dalam Tabel 2.

Tabel 1. Perhitungan *Cosine Similarity*

Term	T(X)	T(Y)
high	2	1
school	2	1
story	1	0
love	1	0
friendship	1	0
student	2	1
face	0	1
had	0	1
complex	0	1
appearance	0	1

Pada Tabel 2, Vektor X dan vektor Y mewakili term "X" dan term "Y" untuk melihat banyaknya nilai yang diperoleh dalam setiap kata yang unik dalam dokumen. Lalu akan masuk ke perhitungan dengan cosine similarity.

a. Vektor X = (2,2,1,1,1,1,0,0,0,0)
b. Vektor Y = (1,1,0,0,0,1,1,1,1,1)

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 10, Oktober 2024
ISSN 2828-2442 (media online)
Hal 2503-2509

Untuk melakukan perhitungan antara term "X"dan term "Y" pada tabel 3.5 akan menggunakan rumus cosine similarity sebagai berikut.

$$\text{Similarity (X, Y)} = \frac{(2x1)+(2x1)+(1x0)+(1x0)+(1x0)+(2x1)+(0x1)+(0x1)+(0x1)+(0x1)}{\sqrt{2^2+2^2+1^2+1^2+1^2+1^2+0^2+0^2+0^2+0^2}\times\sqrt{1^2+1^2+0^2+0^2+0^2+1^2+1^2+1+1^2+1^2}}$$

$$\text{Similarity (X,Y)} = \frac{2+2+0+0+0+2+0+0+0+0}{\sqrt{2+2+1+1+1+1+0+0+0+0}\times\sqrt{1+1+0+0+0+1+1+1+1+1}}$$

$$\text{Similarity (X,Y)} = \frac{6}{\sqrt{8}\times\sqrt{7}}$$

$$\text{Similarity (X,Y)} = \frac{6}{7.48} = 0.80$$

4. IMPLEMENTASI

4.1 *Preprocessing* Data

Preprocessing data yang pertama adalah case folding, yaitu pengubahan kata-kata yang ada dalam data film maupun drama, seperti judul, overview, genre, dan cast. Tujuan dari proses ini yaitu agar tiap kata yang ada dalam data menjadi sama dan setara. Lalu yang kedua adalah tokenizing, dimana proses pemotongan terhadap string input berdasarkan pada setiap kata. Implementasinya dapat dilihat dalam gambar 4.

Gambar 4. *Preprocessing Data*

4.2 Implementasi TF-IDF dan *Filtering* dalam Pyhton

Gambar 5 menunjukan algoritma TF-IDF untuk mengukur dan memberi bobot pada kata-kata yang terdapat dalam data film dan drama dan melakukan filtering pada tiap kata dalam data.

Gambar 5. TF-IDF dan *Filtering*

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 10, Oktober 2024
ISSN 2828-2442 (media online)
Hal 2503-2509

4.3 Implementasi Perhitungan *Cosine Similarity*

Setelah melalui pembobotan dengan TF-IDF, maka dilakukan perhitungan dengan *cosine similarity* agar mendapatkan hasil rekomendasi drama dan juga film yang akurat berdasarkan kata yang telah di proses dalam data. Implementasi *cosine similarity* dalam kode dapat dilihat dalam gambar 6.

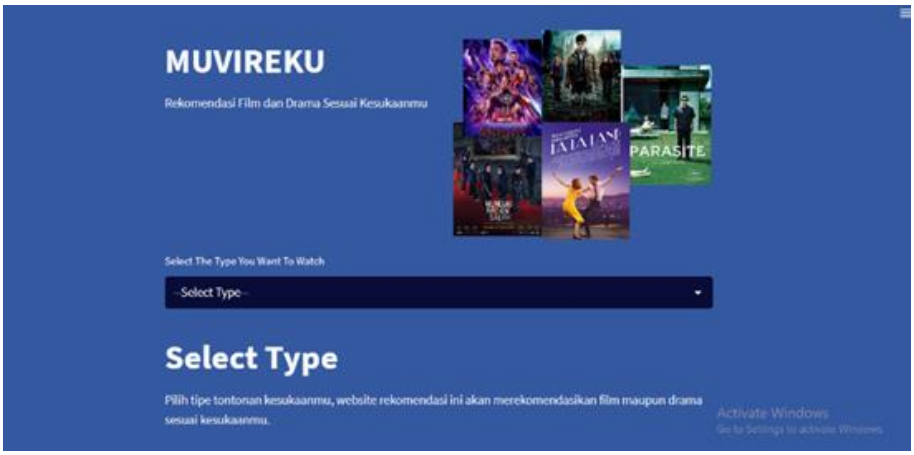
```
43]: # Import linear_kernel
from sklearn.metrics.pairwise import linear_kernel

# Compute the cosine similarity matrix
cosine_sim = linear_kernel(tfidf_matrix, tfidf_matrix)
```

Gambar 6. *Cosine Similarity*

4.4 *User Interface*

a. Tampilan Halaman Utama

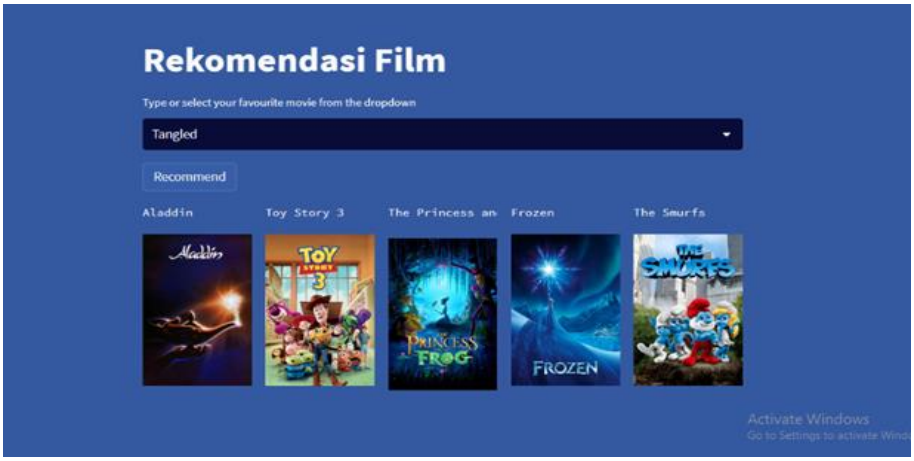
Saat *user* masuk kedalam website, maka tampilan awal akan seperti yang terlihat pada gambar 7. Disana ada *dropdown* untuk memilih tipe tontonan yang diinginkan oleh *user*.



Gambar 7. Tampiilan Halaman Utama

b. Tampilan Rekomendasi

Gambar 8 menunjukan rekomendasi film berdasarkan judul yang di input *user* kedalam sistem, akan keluar 5 rekomendasi film.



Gambar 8. Tampilan Rekomendasi

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 10, Oktober 2024
ISSN 2828-2442 (media online)
Hal 2503-2509

5. KESIMPULAN

Berdasarkan pengujian yang dilakukan pada system rekomendasi, dapat disimpulkan bahwa website rekomendasi dapat berjalan dengan baik dari segi user interface, fungsi fitur yaitu memilih tipe yang akan dicari(drama atau film) lalu memberikan rekomendasi terhadap pengguna berdasarkan preferensi pengguna. Dengan memasukkan judul film maupun drama favorit, maka system akan merekomendasikan item yang relevan dengan yang di input. Pada penelitian selanjutnya diharapkan untuk menambah fitur akses yang akan langsung menuju situs streaming yang akan memudahkan pengguna dalam menonton film di situs resminya.

REFERENCES

Sebastia, L., Garcia, I., Onaindia, E. & Guzman, C. 2009. "e-Tourism: A tourist recommendation and planning application, International Journal on Artificial Intelligence Tools"

Muslimah, N. 2018. "Klasifikasi film berdasarkan sinopsis dengan menggunakan improved K-Nearest Neighbor (K-NN)". (Doctoral dissertation, Universitas Brawijaya).

B. Utomo and Y. Suhari. 2013. "Rekomendasi Film Berbasis Web Pada Bioskop Mini Menggunakan Algoritma Nearest-Neighbor," vol. 5, no. 1.

Hunt, N., dan Gomez-Uribe, C A., 2015, "The Netflix Recommender System: Algorithms, Business Value, and Innovation," ACM Transactions on Management Information Systems, vol. 6, no. 4.

Putra, D. W. T., & Andriani, R. (2019). Unified modelling language (uml) dalam perancangan sistem informasi permohonan pembayaran restitusi sppd. Jurnal Teknoif Teknik Informatika Institut Teknologi Padang, 7(1), 32-39.

Trisno, I. B., & Hari, Y. (2021). Desain dan Analisa Sistem Magang di Prodi Teknik Informasi Universitas Widya Kartika Menggunakan UML. Jurnal Nasional Komputasi dan Teknologi Informasi, 4(6), 490-501.

Alkaff, M., Khatimi, H., & Eriadi, A. (2020). Sistem Rekomendasi Buku Menggunakan Weighted Tree Similarity dan Content Based Filtering. MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput, 20(1), 193-202.

Chiny, M., Chihab, M., Bencharef, O., & Chihab, Y. (2022). Netflix Recommendation System based on TF-IDF and Cosine Similarity Algorithms. no. Bml, 15-20.

Arfisko, H. H., & Wibowo, A. T. (2022). Sistem Rekomendasi Film Menggunakan Metode Hybrid Collaborative Filtering Dan Content-Based Filtering. eProceedings of Engineering, 9(3).

Putri, M. W., Muchayan, A., & Kamisutara, M. (2020). Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF. JOINTECS (Journal of Information Technology and Computer Science), 5(3), 229-236.