

REKOMENDASI DRAMA KOREA MENGGUNAKAN *CONTENT-BASED FILTERING* DENGAN *WEIGHTED TREE SIMILARITY*

Yuricha^{1*}, Irwan Kurnia Phan²

¹Institut Teknologi dan Bisnis Sabda Setia; Jalan Purnama 2 Pontianak; Hp 628561117855

²Universitas Widya Dharma Pontianak; Jalan H.O.S. Cokroaminoto Pontianak; Telp (62561) 731966

Received: 2 Januari 2025

Accepted: 14 Januari 2025

Published: 20 Januari 2025

Keywords:

Korean Drama; Content Based Filtering; Ground Truth; Paradox of Choice; Weighted Tree Similarity;

Correspondent Email:

yuricha@itbss.ac.id

Abstrak. Setiap tahun, ratusan drama Korea dirilis, menghadirkan tantangan bagi penonton untuk memilih tontonan yang sesuai dengan preferensi mereka. Fenomena *Paradox of Choice* menggambarkan bagaimana manusia cenderung bingung saat dihadapkan dengan banyak pilihan dan salah satu cara mengatasinya adalah dengan sistem rekomendasi yang dapat membantu pengguna dalam mempercepat pengambilan keputusan dan mencapai kenyamanan pengguna. Penelitian ini bertujuan membangun sistem rekomendasi drama Korea menggunakan metode *Content Based Filtering* (CBF) dengan algoritma *Weighted Tree Similarity* untuk membantu penonton memilih drama berdasarkan kesamaan konten. Model dikembangkan dengan pembobotan fitur, di mana *screenwriter* memiliki bobot tertinggi (6), diikuti synopsis (2), *actor* (1), dan *director* (1). Evaluasi model dilakukan menggunakan *ground truth* dari rekomendasi MyDramaList dengan metrik *precision*, *recall*, dan *F1-score*. Hasil evaluasi menunjukkan nilai *precision* 0,80 dan *recall* 0,75, dengan *F1-score* sebesar 0,774.

Abstract. Every year, hundreds of Korean dramas are released, posing a challenge for viewers to choose the shows that match their preferences. The *Paradox of Choice* phenomenon illustrates how people tend to feel confused when faced with an overwhelming number of options, and one way to address this issue is through recommendation systems that can help users in expediting decision-making and enhancing their overall experience. This study aims to develop a Korean Drama recommendation system using the *Content-Based Filtering* (CBF) method with the *Weighted Tree Similarity* algorithm to help viewers choose dramas based on content similarity. The model is developed using feature weighting, where screenwriters are assigned the highest weight (6), followed by synopsis (2), actors (1), and directors (1). The model evaluation is conducted using ground truth from MyDramaList, employing precision, recall, and F1-score metrics. The evaluation results indicate a precision of 0.80, recall of 0.75, and an F1-score of 0.774.

1. PENDAHULUAN

Ratusan drama korea dirilis setiap tahunnya, penonton cenderung akan menonton yang sesuai dengan preferensinya baik dari segi genre, synopsis, aktor yang berperan, penulis, bahkan sutradaranya. Hal ini juga

meningkatkan potensi penonton juga dapat mengalami fenomena *Paradox of Choice* [1].

Paradox of Choice merupakan fenomena yang berkaitan dengan kondisi psikologis dimana manusia memiliki kecenderungan semakin bingung ketika dihadapkan dengan banyak pilihan dan menyebabkan manusia sulit

mengambil keputusan [2]. Ketika dihadapkan dengan berbagai tontonan yang ada, manusia cenderung menjadi bingung dan ragu untuk menonton salah satu tontonan, karena takut ketinggalan tontonan lainnya (*FOMO/Fear of Missing Out*).

Rekomendasi menjadi salah satu cara mengatasi *Paradox of Choice*. Berbagai rekomendasi dengan algoritma tertentu dibuat untuk membantu penonton mengambil keputusan tentang drama apa yang akan ditonton serta meningkatkan pengalaman pengguna [3].

Sistem rekomendasi menggunakan *Content-Based Filtering* (CBF) sebelumnya pernah dilakukan untuk memberikan rekomendasi film pada platform streaming dengan hanya memanfaatkan TF-IDF dan *Cosine Similarity* namun tidak ada pembobotan fitur dan rekomendasi yang diberikan masih dibedakan per tipe dan hanya mencari persamaan dari deskripsi film tersebut [4], [5]. Deskripsi film biasanya akan berbentuk banyak format dan tergantung pada kreativitas kru film untuk menarik minat penonton sehingga tidak bisa hanya satu atribut saja yang digunakan.

Rekomendasi film lainnya yang juga menggunakan CBF diuji menggunakan cosine similarity terhadap kueri *single seed* sebesar 0.823254 dan 0.7500556 untuk kueri *multiple seeds*. Rekomendasi menunjukkan bahwa melalui model ini sistem dapat merekomendasikan film dengan tema yang sama dengan menggunakan fitur judul film, genre, dan sinopsis. Ketiga parameter memiliki level TF-IDF yang sama dengan adalah pembobotan term sehingga sangat memungkinkan menghasilkan rekomendasi yang bertumpu pada parameter *genre* saja [6].

Penelitian tentang rekomendasi drama korea pernah dilakukan menggunakan *Hybrid Filtering* [7]–[11][12]. *Hybrid Filtering* memungkinkan untuk menggabungkan CBF dan *Collaborative Filtering*. Begitu juga dengan penelitian lainnya menggunakan *Hybrid Filtering* yang menambahkan variabel lain hasil *crawling* data dari sosial media seperti Twitter/X [13], [14]. Namun untuk dapat menggunakan metode ini, metadata memerlukan atribut yang lebih mendetail dan dataset dengan atribut terbatas tidak akan cocok menggunakan metode ini [15]. Penelitian

tersebut juga tidak memiliki *ground truth* untuk mengevaluasi model yang diuji.

Rekomendasi drama korea berdasarkan kesamaan konten juga pernah diteliti sebelumnya dengan menggunakan 5 (lima) atribut yaitu genre, penulis, sutradara, penayangan, dan aktor utama. Data yang digunakan hanya sebanyak 25 data sampel dengan nilai kemiripan tertinggi sebesar 0.6 [16]. Sampel yang digunakan terlalu sedikit dan wajar jika nilai kemiripan tertinggi tidak lebih dari 70%. Penelitian ini berfokus pada pengembangan prototipe, sehingga hasil rekomendasi menjadi sajian setelah prototipe. Semakin banyak data yang dimasukkan dalam pembelajaran mesin, akan membantu dalam memberikan rekomendasi yang lebih baik kepada pengguna dalam rangka mempercepat pengambilan keputusan [17].

Penelitian terkait rekomendasi menggunakan CBF juga pernah dilakukan untuk merekomendasi berita artikel sebanyak 10 berita teratas [18] dengan *recall* sebesar 80% dan juga pembobotan pada parameter memungkinkan memberikan rekomendasi yang baik bagi pengguna, namun pengujian dengan menggunakan *ground truth* tetap diperlukan untuk mengevaluasi hasil model dengan yang terjadi di lapangan.

Penelitian lainnya menggunakan CBF adalah penelitian sistem merekomendasikan café [19]. Penelitian ini membuktikan bahwa dengan data yang terbatas, sistem mampu memberikan rekomendasi untuk memudahkan pengguna mencari café ataupun minuman di café yang sesuai dengan preferensi sebelumnya. Penambahan parameter dan data akan sangat memungkinkan bagi peneliti untuk memberikan rekomendasi yang lebih efektif. Kualitas dataset sebagai konten yang sangat penting diperhatikan untuk mendapatkan hasil rekomendasi yang berkualitas [20].

CBF memungkinkan untuk memberikan rekomendasi dengan dataset terbatas dan memaksimalkan nilai fitur yang ada dengan pembobotan seperti yang dilakukan pada penelitian untuk mencari skripsi mahasiswa [21]. Penelitian ini menggunakan *Weighted Tree Similarity* sehingga parameter yang digunakan tidak hanya satu tetapi bisa lebih dari satu dan memberikan bobot terhadap tiap parameter untuk meningkatkan hasil rekomendasi. Hasil penelitian menunjukkan

bahwa nilai *precision* berada pada angka 74% dengan *recall* 83%, hasil ini diujikan kembali dengan penerapan dalam memberikan rekomendasi drama korea. Sehingga penelitian dengan CBF untuk rekomendasi film Indonesia juga bisa memiliki nilai presisi yang lebih tinggi dengan adanya pembobotan fitur [22].

Oleh karena itu, penelitian ini berfokus pada membangun rekomendasi drama korea menggunakan CBF dengan *Weighted Tree Similarity* dan mengevaluasi model dengan menggunakan *ground truth* berupa rekomendasi asli dari komunitas drama korea seperti MyDramaList [23]. Model rekomendasi dibangun dengan menggunakan bahasa pemrograman *Python*.

2. TINJAUAN PUSTAKA

Penelitian ini berdasarkan teori-teori pendukung seperti *Content Based filtering*, TF-IDF, *Cosine Similarity*, *Weighted Tree Similarity*, dan *Ground Truth*.

2.1. Content Based Filtering (CBF)

CBF menggunakan pendekatan matematis untuk memberikan rekomendasi dengan salah satu cirinya adalah penggunaan TF-IDF yang memungkinkan konten diolah kata demi kata sehingga konten yang ada dapat dihitung nilai kesamaannya dan membandingkannya dengan item lainnya agar dapat memberikan rekomendasi berdasarkan nilai kesamaan yang tinggi [3].

CBF merupakan salah satu metode yang digunakan untuk memberikan rekomendasi dengan personalisasi (*personalized*). CBF merekomendasikan item yang mirip dengan item yang disukai pengguna sebelumnya. CBF juga mempelajari profil preferensi pengguna baru berdasarkan data dari objek yang telah dinilai oleh pengguna sebelumnya. CBF sebagai algoritma machine learning yang cocok digunakan terutama ketika kurangnya data riwayat interaksi yang cukup untuk dilatihkan menggunakan pembelajaran mesin [24].

Dua informasi penting yang perlu ada untuk menggunakan metode CBF adalah model preferensi pengguna dan riwayat interaksi pengguna. Salah satu kelebihan dari penggunaan CBF adalah tidak menggantungkan pada riwayat data pengguna yang biasanya bersifat sensitif dan memiliki risiko pelanggaran privasi (*Ethical AI*).

Metode CBF sebenarnya tidak hanya dapat dilakukan dengan menggunakan *Cosine Similarity* namun juga dapat menggunakan *Jaccard Similarity*. Namun, pada penelitian yang dilakukan pada aplikasi *Eventhings*, *Cosine Similarity* dapat memberikan nilai *diversity* yang lebih baik dibandingkan *Jaccard Similarity* sebesar 66% [25] sehingga penggunaan *Cosine Similarity* diharapkan akan meningkatkan presisi nilai rekomendasi pada penelitian ini [26].

2.2. TF-IDF

Term Frequency (TF) - *Inverse Document Frequency* (IDF) membuat matriks kata menjadi vektor. TF membandingkan banyaknya suatu kata muncul dengan seluruh jumlah kata tersebut dalam satu dokumen. Sedangkan IDF merupakan proporsi dokumen yang mengandung seluruh kata yang dicari [16]. TF-IDF pada pemrograman python menggunakan `TfidfVectorizer()` yang ada pada library *Sklearn*.

2.3. Cosine Similarity

Cosine Similarity digunakan untuk mengukur kesamaan antara dua vektor dan menentukan apakah kedua vektor menunjuk ke arah yang sama. Semakin besar sudut cosinus antara kedua vektor maka semakin besar nilai *cosinus*-nya [27]. Pada bahasa pemrograman *Python*, *Cosine Similarity* dihitung menggunakan library *Sklearn* untuk mendapatkan nilai *cosinus* antara dua vektor yang disajikan dalam bentuk matriks.

2.4. Weighted Tree Similarity

Weighted Tree Similarity merupakan bagian dari *Cosine Similarity* yang menambahkan dimensi baru dalam analisis konten. *Weighted Tree Similarity* memungkinkan interpretasi yang lebih jelas mengenai suatu konten yang direkomendasikan. *Weighted Tree Similarity* memberikan transparansi dengan atribut fitur yang akan dibobot. Atribut-atribut yang akan digunakan terlebih dahulu didefinisikan dan disesuaikan bobotnya untuk setiap elemen, sehingga hubungi multi-level dan multi-dimensional dalam data dapat ditangkap dengan lebih baik walaupun ada keterbatasan ketersediaan data pengguna [21].

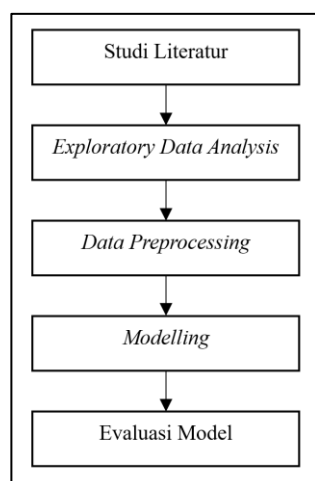
2.5. Ground Truth

Ground truth atau kebenaran dasar merujuk pada data atau informasi yang dianggap benar dan digunakan sebagai acuan untuk mengukur akurasi model rekomendasi [28]. *Ground truth* berupa referensi atau penilaian pengguna yang nyata terhadap item tertentu dan sangat mempengaruhi kinerja sistem rekomendasi. Salah satu isu yang mungkin terjadi pada *ground truth* adalah adanya *preference pollution*.

Preference pollution pada *ground truth* merupakan kondisi di mana data preferensi tercemar oleh faktor eksternal seperti bias budaya, *noise*, dan pengukuran subjektif.

3. METODE PENELITIAN

Metode penelitian yang digunakan pada penelitian ini menggunakan tahapan-tahapan seperti pada Gambar 1.



Gambar 1. Metode Penelitian

3.1. Studi Literatur

Tahapan ini bertujuan untuk memahami konsep dasar sistem rekomendasi, teknik *Content Based Filtering*, serta algoritma *Weighted Tree Similarity*. Kajian dilakukan terhadap penelitian sebelumnya yang relevan, baik dalam domain sistem rekomendasi maupun pendekatan berbasis metadata dalam konteks hiburan. Literatur ini juga mencakup studi tentang dataset yang dapat digunakan untuk pembelajaran mesin seperti Kaggle. Struktur hierarki metadata dari drama Korea yang ada juga harus mampu memberikan informasi rinci untuk memastikan metode yang diusulkan dapat diimplementasikan [29].

3.2. Exploratory Data Analysis

Pada tahap ini, dataset dianalisis dan dieksplorasi lebih jauh setiap variabel-nya secara mendalam menggunakan metode *Univariate Exploratory Data Analysis*. EDA bertujuan untuk memahami pola data, distribusi atribut seperti nama drama, tahun rilis, sinopsis, aktor, rating, dan lainnya serta menemukan keterkaitan antar-atribut. Tahapan ini juga mencakup identifikasi data yang tidak lengkap atau anomali yang dapat memengaruhi hasil rekomendasi. Visualisasi data digunakan untuk memberikan gambaran yang jelas mengenai distribusi metadata pada dataset [30].

3.3. Data Preprocessing

Tahap ini berfokus pada pengolahan data agar siap digunakan dalam proses pemodelan. Tujuan dari tahapan ini adalah mempersiapkan data agar dapat diolah dan memastikan konsistensi, seperti penyalarsan format, pengisian nilai yang hilang (*missing values*), dan normalisasi bobot pada atribut metadata. Selain itu, struktur hierarki pada metadata dibangun untuk memungkinkan penerapan algoritma *Weighted Tree Similarity* secara efektif. Dengan adanya pembobotan fitur, diharapkan nantinya mesin akan mendahulukan atribut dengan bobot yang lebih tinggi baru ke atribut dengan bobot yang lebih rendah.

Dataset yang digunakan sebagai *ground truth* juga akan dipersiapkan agar dapat digunakan untuk mengevaluasi model rekomendasi.

Akhir dari tahapan ini adalah menghasilkan penggabungan dari seluruh data drama, aktor, dan *review* yang telah dipersiapkan sehingga menjadi satu kesatuan data yang utuh dan siap digunakan untuk pemodelan.

3.4. Modelling

Pada tahap ini, metode *Content Based Filtering* menggunakan algoritma *Weighted Tree Similarity* menggunakan *Cosine Similarity*. *Degree of similarity* ini yang akan dimasukkan dalam mesin pembelajaran dan dimasukkan dalam model. Algoritma ini memanfaatkan struktur hierarki metadata dengan pemberian bobot pada setiap tingkat atribut, seperti sinopsis, aktor yang berperan, sutradara, dan penulis naskah. Model ini dikembangkan untuk menghitung kesamaan antar drama berdasarkan preferensi pengguna,

dengan mempertimbangkan bobot yang telah ditentukan.

3.5. Evaluasi Model

Tahap evaluasi dilakukan untuk mengukur performa sistem rekomendasi yang dibangun. Evaluasi mencakup metrik seperti *precision*, *recall*, dan *F1-score* untuk menilai kualitas rekomendasi [31] dengan mengacu pada *ground truth*. Selain itu, dilakukan analisis terhadap skenario *cold start* untuk menguji efektivitas sistem dalam kondisi data terbatas.

4. HASIL DAN PEMBAHASAN

Dataset berupa data drama korea yang dikumpulkan dari tahun 2015 sampai 2023 [23]. Dataset berupa 4 (empat) berkas, yaitu berkas drama korea itu sendiri (*korean_drama.csv*) dengan jumlah data sebanyak 1.752 baris data, berkas *review* drama (*reviews.csv*) dengan baris data sebanyak 10.625, berkas daftar aktor/aktris drama korea (*wiki_actors.csv*) yang memiliki 3.090 baris data, serta berkas rekomendasi dari MyDramaList yang akan digunakan sebagai *ground truth* memiliki 1.753 baris data. Keseluruhan berkas data ditampung pada variabel objek masing-masing untuk selanjutnya masuk ke tahapan *Exploratory Data Analysis* (EDA).

Data *korean_drama.csv* memiliki 17 kolom yang berisi informasi seperti yang ditampilkan pada Gambar 2.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1752 entries, 0 to 1751
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   kdrama_id             1752 non-null   object
1   drama_name           1752 non-null   object
2   year                 1752 non-null   int64
3   director            1036 non-null   object
4   screenwriter         959 non-null    object
5   country              1752 non-null   object
6   type                 1752 non-null   object
7   tot_eps              1752 non-null   int64
8   duration             1728 non-null   float64
9   start_dt             1752 non-null   object
10  end_dt               1752 non-null   object
11  aired_on             1520 non-null   object
12  org_net              1344 non-null   object
13  content_rt           1752 non-null   object
14  synopsis             1584 non-null   object
15  rank                 1752 non-null   int64
16  pop                  1752 non-null   int64
dtypes: float64(1), int64(4), object(12)
```

Gambar 2. EDA Dataset Drama Korea

Dari Gambar 2 dapat dilihat bahwa dari total 1.752 baris data, kolom *director* (sutradara), *screenwriter* (penulis naskah), *aired_on* (jadwal hari penayangan), *org_net* (tempat penayangan), dan *synopsis* (sinopsis drama)

memiliki *missing value*. Hanya 1 kolom yang bertipe *float64*, 4 kolom bertipe *int64* (*year*, *tot_eps*, *rank*, dan *pop*), serta lainnya bertipe *object*.

Data pada *wiki_actors.csv* memiliki 5 (lima) kolom yang terdiri atas *actor_id* (ID aktor dari MyDramaList), *actor_name* (nama aktor), *drama_name* (nama drama yang pernah dimainkan), *character_name* (nama karakter aktor di drama yang dimainkan), dan *role* (tipe peran aktor di drama tersebut, apakah pemeran utama / *main lead*, atau pemeran pendukung / *supporting role*). Data aktor ini tidak memiliki *missing value* di semua kolomnya. Dari data aktor ini, terlihat sebanyak 1.643 drama korea yang masuk dalam list drama yang pernah dimainkan oleh aktor.

Sedangkan pada data *review* drama korea memiliki 10 kolom yang terdiri atas kolom-kolom beserta rinciannya seperti pada Gambar 3.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10625 entries, 0 to 10624
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   user_id              10625 non-null  object
1   title                10625 non-null  object
2   story_score          10625 non-null  float64
3   acting_cast_score    10625 non-null  float64
4   music_score          10625 non-null  float64
5   rewatch_value_score  10625 non-null  float64
6   overall_score        10625 non-null  float64
7   review_text          10619 non-null  object
8   ep_watched           10625 non-null  object
9   n_helpful            10625 non-null  int64
dtypes: float64(5), int64(1), object(4)
memory usage: 830.2+ KB
```

Gambar 3. EDA Dataset Review Drama

Dari Gambar 3 terlihat bahwa hanya kolom *review_text* (berisi *review* dalam bentuk deskripsi kata yang diberikan pengguna terhadap drama) yang memiliki *missing value* sebanyak 6 baris data, sedangkan lainnya terisi.

Dari analisis yang ada, kolom *score* memiliki rentang nilai dari 0-10. Dari total 10.625 baris data yang ada pada data *review*, jumlah drama yang masuk dalam data *review* ada sebanyak 1.279 baris data dan di-review oleh sebanyak 4.562 baris pengguna unik.

Tahapan terakhir dari EDA adalah menganalisis dan mengeksplorasi dataset *ground truth* pada berkas *recommendations.csv*. Dataset hanya memiliki 2 (dua) kolom, yaitu *kdrama_name* (nama drama korea) dan *recommendations* (berupa list drama yang direkomendasikan oleh komunitas MyDramaList). Contoh data pada berkas tersebut dapat dilihat pada Gambar 4.

	kdrama_name	recommendations
0	Mask Girl	['Sympathy for Lady Vengeance', 'True Beauty'...
1	Sing My Crush	['About Youth', 'Wish You: Your Melody From My...'...
2	D.P. Season 2	['Designated Survivor: 60 Days', 'D.P.', 'Weak...'...
3	Not Others	['You're All Surrounded', 'Hometown Cha-Cha-Ch...'...
4	Shadow Detective Season 2	['WATCHER', 'Stranger', 'Stranger Season 2', '...'...

Gambar 4. Sampel Data Ground Truth

Terlihat pada Gambar 4, setiap drama memiliki rekomendasi berupa list yang berupa nama-nama drama korea lainnya yang nantinya akan digunakan untuk mengevaluasi model. Dari total 1.753 baris data yang didapat sebagai ground truth, ada sebanyak 573 baris drama korea yang tidak memiliki daftar rekomendasi sehingga harus diamputasi dari data.

Setelah menganalisis keseluruhan dataset yang akan digunakan, tahapan berikutnya adalah data preprocessing. Hal pertama yang akan dilakukan adalah menggabungkan data aktor dan data drama korea sehingga didapatkan data drama korea beserta aktor yang memerankannya. Agar tidak terjadi duplikasi data, dilakukan grouping data aktor sehingga kesemua nama aktor yang memainkan drama tersebut dapat digabungkan dalam satu kolom membentuk list.

Dari gabungan data aktor dan drama membentuk variabel baru bernama dramas dan akan digabungkan kemudian dengan data review. Satu drama dapat memiliki lebih dari satu review, sehingga keseluruhan nilai score digabungkan menjadi satu dengan menjumlahkan keseluruhan nilai. Seluruh penggabungan yang telah diproses ini akan ditampung pada variabel baru yang dinamai all_drama.

Setelah data telah digabungkan, berbagai dataset digabungkan untuk membentuk dataset komprehensif yang mencakup atribut seperti judul drama, aktor, sutradara, penulis naskah, sinopsis, dan rating. Nilai yang hilang pada atribut kritis seperti "sutradara" dan "penulis naskah" dengan persentase di bawah 20% diisi dengan nilai "Unknown". Sementara itu, data dengan atribut penting yang kosong seperti "sinopsis" dan "aktor" dihapus untuk menjaga integritas dataset. Pada tahapan ini, data yang kosong dan duplikat lainnya telah ditangani sehingga dapat dilanjutkan ke tahap berikutnya.

Data yang telah digabungkan memiliki jumlah baris data akhir terdiri dari 835 data drama dengan tujuh atribut utama yang siap digunakan dalam pemodelan. 5 sampel data yang telah berhasil diolah dapat dilihat seperti pada Gambar 5.

	name	actor	director	screenwriter	rank	overall_score	synopsis
0	100 Days My Prince	['Doh Kyung So', 'Yoon Se Hye', 'Jo Sung Hye'...	['Lee Jong Jai', 'Yoon Sang Wook'...	['Park Ji Suk']	882	6.0	Lee Yul, Crown Prince of Joseon, is a perfect...
1	19 Again	['Kim Ha Neul', 'Lee Da Hyeon', 'Huh Jung Eui'...	['Ha Byung Hoon']	['Yoon Da Yeon', 'Kim Eun Bin', 'Choi Yi Hyun'...	224	7.0	After nearly twenty years of marriage, Jung Du...
2	20th Century Boy and Girl	['Ahn Se Hui', 'Yoon Ye Seul', 'Kim Ji Suk', 'Lee...'...	['Lee Dong Hyeon']	['Lee Seon Hye']	3014	7.0	A top actress/singer who has an innocent soul...
3	365: Repeat the Year	['Shin Seung Gyeon', 'Kim Jae Sook', 'Lee Joon Hyu'...	['Kim Kyung Hae']	['Lee Seo Yoon', 'Lee Soo Kyung']	413	7.5	A story where ten people get the chance to go...
4	38 Task Force	['Choi Sin Young', 'Yu Gye Ri', 'Noo Jae Hee'...	['Yoon Dong Hyeon']	['Yoon Dong Hyeon']	1686	9.5	A special task force will risk cops and criminals...
600	Your Honor	['Heo Ji Won', 'Kwon Na Ra', 'Lee Yoo Young'...	['Seo Sang Chul']	['Chun Sung Il']	1794	9.0	There are identical twins named Han Soo Ho and...
601	Youth of May	['Gwon Sun Hee', 'Go Min Si', 'Lee Do Hyun'...	['Song Min Yeop', 'Lee Dae Kyung']	['Lee Kang']	165	9.5	Yoon Tae becomes the pride of Gyeongju when he i...
602	Yumi's Cells	['Ahn Se Hyeon', 'Joo Jung Hyeon', 'Kim Go Eun'...	['Lee Sang Yeon']	['Song Jae Jung', 'Kim Yoon Joo', 'Kim Kyung Ri'...	630	9.0	Controlled by a complex network of cells, each...
603	Yumi's Cells Season 2	['Ahn Se Hyeon', 'Kim Go Eun', 'Yoon Seung Hyeon'...	['Lee Sang Yeon']	['Song Jae Jung', 'Kim Yoon Joo', 'Kim Kyung Ri'...	615	1.0	After suffering a painful breakup, Yu Mi's sta...
604	Zombie Detective	['Ahn Se Hui', 'Choi Jin Hee', 'Yoojung Ba Hee'...	['Shin Jae Hyun']	['Baek Eun Joo']	1312	6.5	A zombie and a waker team up to solve crimes...

Gambar 5. Sampel Data Hasil Preprocessing

Gambar 5 menunjukkan bahwa data yang akan digunakan untuk rekomendasi drama korea sebanyak 835 baris data dan 7 (tujuh) kolom yang terdiri atas name (nama drama korea), actor (berisi list aktor yang berperan dalam drama tersebut), director (sutradara drama), screenwriter (penulis naskah), rank (ranking drama di MyDramaList), overall_score (penilaian drama korea oleh komunitas MyDramaList), dan synopsis (sinopsis drama korea tersebut).

Selanjutnya, atribut-atribut ini distandarisasi dan digabungkan menjadi fitur gabungan (combined features) untuk memfasilitasi perhitungan kesamaan berbasis fitur dengan algoritma Weighted Tree Similarity. Setiap drama direpresentasikan sebagai vektor dengan bobot tertentu pada fitur: screenwriter (bobot 6), synopsis (bobot 2), actor (bobot 1), dan director (bobot 1). Jumlah pembobotan pada fitur menggunakan trial-and-error untuk memastikan bahwa komunitas merekomendasikan drama berdasarkan penulis naskah dan juga sinopsis yang berperan penting sehingga pendekatan pembobotan fitur dirancang untuk memprioritaskan narasi dan pemain dalam menentukan rekomendasi. Hasil dari TF-IDF membentuk matrix (835, 8532).

Tahapan berikutnya adalah menghitung derajat kesamaan menggunakan Cosine Similarity dan didapatkan hasil seperti pada Gambar 6.


```
[ ] #menghitung cosine similarity pada matrix tf-idf
cosine_sim = cosine_similarity(tfidf_matrix)
cosine_sim

array([[1.0, 0.06332223, 0.14392764, ..., 0.07204165, 0.0695508,
0.0624844 ],
[0.06332223, 1.0, 0.11718497, ..., 0.10268973, 0.11369643,
0.1133178 ],
[0.14392764, 0.11718497, 1.0, ..., 0.09518878, 0.09754618,
0.09374298],
...,
[0.07204165, 0.10268973, 0.09518878, ..., 1.0, 0.72096063,
0.08562047],
[0.0695508, 0.11369643, 0.09754618, ..., 0.72096063, 1.0,
0.07993683],
[0.0624844, 0.1133178, 0.09374298, ..., 0.08562047, 0.07993683,
1.0]])
```

Gambar 6. Hasil Cosine Similarity

Dari Gambar 6 terlihat beberapa nilai menghasilkan nilai 1.0 dan sisanya memiliki hasil yang *diversity*. Keseluruhannya akan disimpan dalam satu dataframe dan akan dilanjutkan dengan pembuatan fungsi untuk memanggil model. Fungsi dikemas dalam pemanggilan kode sebagai berikut:

```
def recommend_drama(title,
cosine_sim=cosine_sim,
data=data, actor_weight=0.1,
director_weight=0.1,
screenwriter_weight=0.6,
synopsis_weight=0.2):
```

Fungsi ini yang dipanggil untuk mendapatkan hasil rekomendasi dari model yang ada. Contohnya target drama korea berjudul *Hospital Playlist* memiliki detail informasi seperti pada Tabel 1.

Tabel 1. Informasi Rinci Target Drama “Hospital Playlist”

Actor	Director	Screenwriter	Combined Features
Jeon Mi Do, Jo Jung Suk, Jung Kyung Ho, Yoon Seok	Shin Won Ho	Lee Woo Jung	Jeon Mi Do' 'Jo Jung Suk' 'Jung Kyung Ho' 'Ki...

Hasil rekomendasi dari target drama “Hospital Playlist” dapat dilihat pada Tabel 2.

Tabel 2. Hasil Rekomendasi dari Target Drama “Hospital Playlist”

Rekomendasi Drama	Actor	Director	Screenwriter	Similarity Degree
Hospital Playlist Season 2	Jeon Mi Do, Jo Jung Suk, Jung Kyung Ho, Yoon Seok	Shin Won Ho	Lee Woo Jung	0.779182
Reply 1988	Go Kyung Pyo, Lee Dong Gwi, Lee Hye Ri, Yoon Seok	Shin Won Ho, Yoo Hak Chan		0.191368
Prison Playbook	Im Hwa Young, Jung Kyung Ho	Shin Won Ho	Jung Bo Hoon, Lee Woo Jung	0.310205

Hospital Playlist Season 2	Jeon Mi Do, Jo Jung Suk, Jung Kyung Ho, Yoon Seok	Shin Won Ho	Lee Woo Jung	0.779182
Reply 1988	Go Kyung Pyo, Lee Dong Gwi, Lee Hye Ri, Yoon Seok	Shin Won Ho, Yoo Hak Chan	Lee Woo Jung	0.191368
Prison Playbook	Im Hwa Young, Jung Kyung Ho	Shin Won Ho	Jung Bo Hoon, Lee Woo Jung	0.310205

Dengan memperhatikan Tabel 1 yang berisi informasi rinci dari drama “Hospital Playlist” dan Tabel 2 berupa hasil rekomendasinya dapat dilihat bahwa drama korea yang direkomendasikan memprioritaskan kolom *Screenwriter* dimana memiliki kesamaan dengan drama target yaitu Lee Woo Jung, disusul dengan *director* dan *actor* yang sama. Walaupun fitur judul drama tidak dibobotkan, namun karena adanya nilai kesamaan *synopsis* yang tinggi dan juga *screenwriter* yang sama membuat sistem dapat memberikan

rekomendasi lanjutan bagi pengguna yang menyukai drama *Hospital Playlist*.

Selanjutnya adalah melakukan evaluasi model dengan menggunakan *ground truth*. *Ground truth* dari drama *Hospital Playlist* menurut MyDramaList dapat dilihat pada Tabel 3.

Tabel 3. *Ground truth* drama *Hospital Playlist*

actor	director	screen writer	synopsis
A Poem a Day	['Jang Dong Yoon', 'Kim Jae Bum', 'Lee Chae Yo...']	['Han Sang Jae']	The story of lives of people who work at the h...
Dr. Romantic	['Choi Jin Ho', 'Han Seok Kyu', 'Jang Hyuk Jin...']	['Yoo In Shik', 'Park Soo Jin']	Kim Sa Bu was once a famous surgeon at the pea...
Hospital Playlist Season 2	['Jeon Mi Do', 'Jo Jung Suk', 'Jung Kyung Ho',...']	['Shin Won Ho']	Everyday is extraordinary for five doctors and...
Live	['Bae Jong Ok', 'Bae Sung Woo', 'Jang Hyun Sun...']	['Kim Kyu Tae', 'Myung Hyun Woo']	The drama revolves around the incidents that o...
Prison Playbook	['Im Hwa Young', 'Jung Kyung Ho', 'Kim Kyung N...']	['Shin Won Ho']	Kim Je Hyuk, a famous baseball player, is arre...
Reply 1988	['Go Kyung Pyo', 'Lee Dong Hwi', 'Lee Hye Ri',...']	['Shin Won Ho', 'Yoo Hak Chan']	Five childhood friends, who all live in the sa...

Dari Tabel 3 terlihat bahwa model rekomendasi berhasil memberikan 3 nama drama yang sama dengan rekomendasi dari MyDramaList, yaitu *Hospital Playlist Season 2*, *Prison Playbook*,

dan *Reply 1988* yang artinya 3 rekomendasi memiliki *true positivity* dengan *ground truth*.

Tahapan berikutnya adalah mengevaluasi model. Evaluasi dilakukan menggunakan metrik *precision*, *recall*, *F1-score*, dan akurasi. Hasil evaluasi menunjukkan:

Precision: Sistem mencatat *precision* sebesar 0,80, menunjukkan bahwa hanya 80% rekomendasi yang benar-benar relevan berdasarkan *ground truth*. Hal ini mengindikasikan adanya beberapa drama yang direkomendasikan tetapi tidak sesuai.

Recall: Nilai *recall* mencapai 0,75, menunjukkan bahwa semua drama yang relevan berhasil direkomendasikan oleh sistem, dengan minim jumlah *false negatives*.

Dari nilai *precision* dan *recall* menunjukkan bahwa sistem cukup baik dalam mengidentifikasi item relevan, tetapi ada beberapa item relevan yang terlewatkan.

F1-score: *F1-score* sebesar 0,774 menunjukkan performa yang cukup baik meski *precision* masih perlu ditingkatkan.

Akurasi: Nilai akurasi sistem sebesar 0,75 mengindikasikan bahwa tingkat *true positivity* yang harus ditingkatkan lagi.

Evaluasi mengungkapkan bahwa meskipun sistem memiliki tingkat *recall* yang lebih rendah dibandingkan nilai *precision*. Walau terlihat baik, namun nilai *precision* masih dapat ditingkatkan dan peneliti menyadari bahwa dengan tidak adanya atribut genre cukup mempengaruhi nilai rekomendasi. Selain itu, kualitas dari *ground truth* juga perlu diperhatikan karena berpotensi mengalami *preference pollution* [28].

5. KESIMPULAN

- Penelitian ini menunjukkan potensi pendekatan *Content Based Filtering* dengan menggunakan algoritma *Weighted Tree Similarity* digunakan untuk merekomendasikan drama korea menggunakan bobot tertentu pada fitur: *screenwriter* (bobot 6), *synopsis* (bobot 2), *actor* (bobot 1), dan *director* (bobot 1) memiliki evaluasi model dengan nilai *precision* yang lebih tinggi dibandingkan nilai *recall*-nya. Hal ini menunjukkan bahwa model yang dibuat memiliki performansi yang baik, namun perlu untuk

- mendapatkan *dataset* dengan metadata yang lebih lengkap untuk meningkatkan nilai *precision* yang singkatnya perlu adanya pengayaan dataset.
- b. Kualitas *ground truth* juga akan menjadi penentu yang sangat penting terutama dalam hal kualitas dan akan lebih baik jika tidak mengalami *preference pollution*.
 - c. Hasil rekomendasi yang lebih maksimal dapat dikombinasikan menjadi hybrid filtering dengan tetap menggunakan algoritma *Weighted Tree Similarity*.

DAFTAR PUSTAKA

- [1] Kosim and R. Prihandi, "Sistem Rekomendasi Menu Minuman dengan Metode Content-Based Filtering Berbasis Android pada Mubtada Kopi," *J. Comput. Sci. Artif. Intell.*, vol. 1, no. 1, pp. 43–69, 2024.
- [2] M. Horta Ribeiro, V. Veselovsky, and R. West, "The Amplification Paradox in Recommender Systems," in *Proceedings of the Seventeenth International AAAI Conference on Web and Social Media (ICWSM 2023)*, 2023, pp. 1138–1142. [Online]. Available: www.aaai.org
- [3] T. Ridwansyah, B. Subartini, and S. Sylviani, "Penerapan Metode Content-Based Filtering pada Sistem Rekomendasi," *Math. Sci. Appl. J.*, vol. 4, no. 2, pp. 70–77, Apr. 2024, doi: 10.22437/msa.v4i2.32136.
- [4] J. M. Azri Saputra, L. M. Huizen, and D. B. Arianto, "Sistem Rekomendasi Film pada Platform Streaming Menggunakan Metode Content-Based Filtering," *J. Transform.*, vol. 22, no. 1, p. 10, Aug. 2024, doi: 10.26623/transformatika.v22i1.7041.
- [5] S. Rakesh, "Movie Recommendation System Using Content Based Filtering," *Al-Bahir J. Eng. Pure Sci.*, vol. 4, no. 1, pp. 63–70, Dec. 2023, doi: 10.55810/2313-0083.1043.
- [6] M. Fajriansyah, P. P. Adikara, and A. W. Widodo, "Sistem Rekomendasi Film Menggunakan Content Based Filtering," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 6, pp. 2188–2199, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] M. R. Sugiri, V. C. Mawardi, and J. Hendryli, "Implementing Hybrid Filtering on Korean Drama Recommendation Through K Nearest Neighbor Algorithm," *Indones. J. Multidisciplinary Sci.*, vol. 1, no. 4, pp. 380–393, 2022.
- [8] I. H. Arsyntania, E. B. Setiawan, and I. Kurniawan, "Movie Recommender System with Cascade Hybrid Filtering Using Convolutional Neural Network," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 10, no. 2, pp. 188–200, 2024, doi: 10.26555/jiteki.v9i4.28146.
- [9] P. Sharma and L. Yadav, "Movie Recommendation System Using Item Based Collaborative Filtering," *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 8, no. 4, Jul. 2020, doi: 10.21276/ijrcst.2020.8.4.2.
- [10] A. Nilla and E. B. Setiawan, "Film Recommendation System Using Content-Based Filtering and the Convolutional Neural Network (CNN) Classification Methods," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 10, no. 1, p. 17, Feb. 2024, doi: 10.26555/jiteki.v9i4.28113.
- [11] Y. V. L. Jaja, B. Susanto, and L. R. Sasongko, "Penerapan Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Data MovieLens," *d'Cartesia N J. Mat. dan Apl.*, vol. 9, no. 2, pp. 73–83, 2020, [Online]. Available: <https://ejournal.unsrat.ac.id/index.php/decartesian>
- [12] W. Kristianto, D. Erny Herwindiati, and J. Hendryli, "Sistem Rekomendasi Drama Korea Menggunakan Metode User-Based Collaborative Filtering," *J. Ilmu Komput. dan Sist. Inf.*, pp. 36–40, 2021.
- [13] M. T. Muhadzdzib Ramadhan and E. B. Setiawan, "Netflix Movie Recommendation System Using Collaborative Filtering With K-Means Clustering Method on Twitter," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 4, p. 2056, Oct. 2022, doi: 10.30865/mib.v6i4.4571.
- [14] A. Fazira Ansshory and E. Budi Setiawan, "Social Media (Twitter) Based Movie Recommendation System On Disney+ With Hybrid Filtering Using Neighbor's K-Nearest Method," *JINAV J. Inf. Vis.*, vol. 4, no. 2, pp. 2746–1440, 2023, doi: 10.35877/454RI.jinav1954.
- [15] C. Christakou, L. Lefakis, S. Vrettos, and A. Stafylopatis, "A Movie Recommender System Based on Semi-supervised Clustering," in *2005 International Conference on Computational Intelligence for Modelling, Control, and Automation, and International Conference on Intelligent Agents, Web Technologies, and Internet Commerce (CIMCA-IAWTIC'05)*, 2005. [Online]. Available: www.imdb.com

- [16] J. Safitri, V. Atina, and N. A. Sudiby, "Rancang bangun sistem rekomendasi pemilihan drama korea dengan metode content-based filtering," *INFOTECH J. Inform. Teknol.*, vol. 5, no. 2, pp. 175–189, 2024, doi: 10.37373/infotech.v5i2.1235.
- [17] H. Oktaviana and N. Latifah, "Enhancing Access and User Experience in K-Drama Consumption: Development and Evaluation of The 'K-Drama Universe' Application," *J. Multiling.*, vol. 4, no. 3, pp. 8–21, 2024.
- [18] A. A. Huda, R. Fajarudin, and A. Hadinegoro, "Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2511.
- [19] A. W. Wicaksono, A. N. Rohman, and A. D. Hartanto, "Café Recommendation Using the Content-Based Filtering Method," *J. Inf. Syst. Informatics*, vol. 6, no. 3, pp. 1598–1615, Sep. 2024, doi: 10.51519/journalisi.v6i3.813.
- [20] Falahah and R. Komalasari, "Model Rekomendasi Makanan Menggunakan Content-Based dan Collaborative Filtering," in *Prosiding Seminar Nasional SISFOTEK*, 2024, pp. 673–678.
- [21] N. Matondang, Y. V. Via, and F. A. Akbar, "Implementasi Algoritma Weighted Tree Similarity dan Content-based Filtering dalam Pencarian Skripsi," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2554–2563, Aug. 2024, doi: 10.23960/jitet.v12i3.4807.
- [22] A. Zakharia, A. D. Ulhaq, A. B. Suryono, N. C. Nugroho, D. F. Hafith, and N. D. A. Gusmao, "Sistem Rekomendasi Film Indonesia Menggunakan Metode Content-Based Filtering," *Log. J. Ilmu Komput. dan Pendidik.*, vol. 2, no. 4, pp. 671–678, 2024, [Online]. Available: <https://journal.mediapublikasi.id/index.php/logic>
- [23] C. Charuchinda, "Korean Drama from 2015-2023 with Actors & Reviews," *Kaggle*, 2024. <https://www.kaggle.com/datasets/chanoncharuchinda/korean-drama-2015-23-actor-and-reviewmydramalist> (accessed Dec. 12, 2024).
- [24] A. I. Putra and R. R. Santika, "Implementasi Machine Learning dalam Penentuan Rekomendasi Musik dengan Metode Content-Based Filtering," *Edumatic J. Pendidik. Inform.*, vol. 4, no. 1, pp. 121–130, Jun. 2020, doi: 10.29408/edumatic.v4i1.2162.
- [25] A. A. Ristias, E. D. Wahyuni, and S. F. A. Wati, "Komparasi Kinerja Metode Cosine dan Jaccard Similarity dalam Content-Based Recommendation Systems (CBRS) pada Aplikasi Eventhings," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2304–2313, Aug. 2024, doi: 10.23960/jitet.v12i3.4738.
- [26] N. Noor Kamala Sari, R. Priskila, and P. Bagus Adidyana Anugrah Putra, "Implementasi Content-Based Filtering Menggunakan TF-IDF dan Cosine Similarity untuk Sistem Rekomendasi Resep Masakan," *J. Teknol. Inf.*, vol. 18, no. 1, 2024, doi: 10.47111/JTI.
- [27] A. H. J. P. Juni Permana and Agung Toto Wibowo, "Movie Recommendation System Based on Synopsis Using Content-Based Filtering with TF-IDF and Cosine Similarity," *Int. J. Inf. Commun. Technol.*, vol. 9, no. 2, pp. 1–14, Dec. 2023, doi: 10.21108/ijoiect.v9i2.747.
- [28] G. Adomavicius, J. C. Bockstedt, S. P. Curley, and J. Zhang, "Recommender systems, ground truth, and preference pollution," *AI Mag.*, vol. 43, no. 2, pp. 177–189, Jun. 2022, doi: 10.1002/aaai.12055.
- [29] V. Sandrya, Wasino, and D. Arisandi, "Sistem Rekomendasi Film Menggunakan Metode Multiple Attribute Utility Theory," *Comput. J. Comput. Sci. Inf. Syst.*, vol. 6, no. 1, pp. 19–30, 2022.
- [30] V. Rolanda, T. S. Gunawan, and Wanayumini, "Content-Based Filtering Recommendation System Using Categories Search Engine," *Int. J. Res. Vocat. Stud.*, vol. 2, no. 4, pp. 120–125, Jan. 2023, doi: 10.53893/ijrvocas.v2i4.177.
- [31] W. G. S. Parwita, "Pengujian Akurasi Sistem Rekomendasi Berbasis Content-Based Filtering," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 14, no. 1, p. 27, Feb. 2019, doi: 10.30872/jim.v14i1.1272.