Sistem Rekomendasi Drama Korea Berdasarkan Prediksi *Rating* dan Kemiripan Sinopsis Menggunakan *Singular Value Decomposition* dan *Term Frequesncy - Inverse Document Frequency*

Laila Novia Sari[1], Anita Qoiriah[2]

[1,2] Program Studi S1 Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya
[1]laila.20013@mhs.unesa.ac.id
[2]anitaqoiriah@unesa.ac.id

***Abstrak*— Pengaruh budaya Korea semakin menyebar luas di Indonesia, menyebabkan banyak drama korea baru dirilis. Hal ini membuat pecinta drama korea bingung dalam memilih drama korea yang ingin mereka tonton. Sehingga dibutuhkan sistem rekomendasi drama korea berdasarkan prediksi *rating* dan kemiripan sinopsis. Pecinta drama korea mayoritas memilih drama korea yang populer, dan *rating* menjadi salah satu pengaruh sebuah drama korea dikatakan populer. Atribut sinopsis untuk memudahkan pecinta drama korea mengetahui alur cerita drama korea. *Singular Value Decomposition* digunakan untuk membuat model prediksi *rating*. *Term Frequency-Inverse Document Frequency* digunakan untuk memberikan bobot kata pada sinopsis dan deskripsi singkat, kemudian dihitung kemiripannya dengan *Cosine Similarity*. Model prediksi *rating* dilatih dan diuji menggunakan lima skenario pembagian data (50:50, 60:40, 70:30, 80:20, 90:10), menghasilkan *RMSE* terendah pada skenario 90:10 sebesar 1.493423. Kemiripan sinopsis dicari dengan lima skenario deskripsi singkat yang berbeda, dan dihasilkan nilai kemiripan tertinggi yaitu 53%. Sistem rekomendasi yang dibuat bisa memberikan rekomendasi drama korea yang sesuai dan bisa diurutkan berdasarkan kemiripan sinopsis dan prediksi *rating*.**

***Kata Kunci*— Sistem Rekomendasi, Drama Korea, Prediksi *Rating*, Kemiripan Sinopsis. *SVD, TF-IDF*.**

I. PENDAHULUAN

Semakin berkembangnya media hiburan dan dampak dari *Korean Wave* yang semakin menyebar luas di Indonesia. Membuat banyak remaja di Indonesia menyukai *K*-Drama, hal ini karena konten yang disajikan unik dan berbeda sehingga memiliki ketertarikan tersendiri [1]. Drama korea atau *K*-Drama umumnya dirilis 1-2 episode tiap minggunya. Namun, karena banyaknya kanal TV dan produser yang memproduksi drama korea, sehingga memungkinkan hampir setiap bulan terdapat drama korea baru yang dirilis. Masalah ini menyebabkan para pecinta drama korea bingung dalam memilih drama korea yang ingin mereka tonton. Sehingga dibutuhkan sebuah sistem yang bisa memberikan rekomendasi secara akurat untuk membantu dalam pemilihan drama korea yang ingin ditonton, sistem yang dimaksud biasa disebut dengan sistem rekomendasi. Sistem rekomendasi adalah sistem yang berfungsi untuk membantu pengguna dalam menemukan dan memilih produk atau jasa yang sesuai dengan kriteria atau ketentuan tertentu dari masing-masing pengguna [2].

Singular Value Decomposition (*SVD*) merupakan salah satu teknik *Matrix Factorization* yang sering digunakan untuk membuat sistem rekomendasi. *SVD* merupakan metode yang merubah matriks *A* (interaksi *user-item*) berdimensi $m \; x \; n$ menjadi perkalian tiga matriks dengan dimensi yang lebih kecil. *Singular Value Decomposition* (*SVD*) menghasilkan nilai MAE dan *RMSE* lebih baik karena menghasilkan prediksi *rating* yang mendekati *rating* aslinya [3].

Dalam memilih drama korea yang ingin ditonton, atribut yang sering digunakan dalam menentukan pilihan drama korea adalah *rating* dan sinopsis. Pecinta drama korea terutama masyarakat Indonesia mayoritas memilih drama korea yang sedang populer karena mengikuti *tren* yang ada, *rating* menjadi salah satu pengaruh sebuah drama korea dikatakan sebagai drama yang populer, semakin tinggi nilai *rating*, semakin bagus atau populer drama tersebut [4]. Berdasarkan beberapa situs *database* drama atau film seperti IMDb, MyDramaList, dan AsianWiki, atribut pertama yang ditampilkan pada tiap drama adalah data *rating*, dengan anggapan bahwa *rating* mempengaruhi user dalam memilih drama mana yang ingin ditonton [5]. *Rating* adalah sebuah satuan yang digunakan untuk mengukur performa suatu *item* (misal film, drama, TV, produk, dan sebagainya). *Item* dengan *rating* yang tinggi diasumsikan sebagai *item* yang bagus atau populer. *Rating* memiliki pengaruh besar bagi seorang user dalam memilih film atau drama, karena sebagian besar dapat mencerminkan preferensi mereka terhadap film atau drama [6].

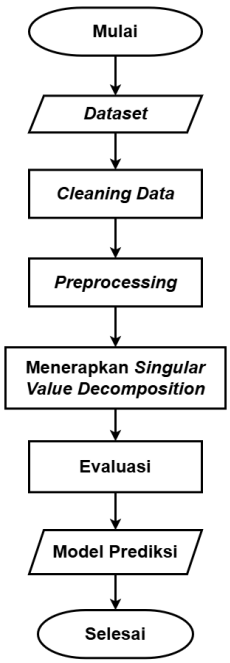
Selain *rating* atribut lain yang menjadi kecenderungan *user* dalam memilih *item* yang ingin ditonton adalah sinopsis atau alur cerita. Sinopsis berisi penjelasan singkat mengenai alur cerita sebuah drama atau film yang bisa membantu pengguna untuk mengetahui garis besar dari drama atau film tersebut [7]. Sinopsis bisa digunakan sebagai parameter dalam mencari kemiripan drama yang telah ditonton dengan drama lainnya [8]. Kemiripan dari sinopsis dapat dicari dengan memberikan bobot kata menggunakan perhitungan *Term Frequency – Inverse Document Frequency* (*TF-IDF*). Metode ini digunakan untuk mengetahui seberapa penting sebuah kata pada suatu dokumen.

Semakin sering suatu kata muncul pada suatu dokumen, semakin besar pula bobot suatu kata. Setelah melakukan pembobotan kata, untuk mencari tingkat kemiripan kata kunci dengan suatu dokumen dapat dihitung menggunakan *cosine similarity*. [9].

Pada penelitian terdahulu menggunakan dua metode *Matrix Factorization* yaitu *Non-negative Matrix Factorization* dan *Probabilistic Matrix Factorization*. Hasil yang diperoleh, diketahui bahwa *NNMF* lebih baik dari pada *PMF*. Pada penelitian selanjutnya diharapkan bisa menggunakan metode *Matrix Factorization* yang lain dan mengkombinasikannya dengan algoritma lain, supaya bisa membuat sistem rekomendasi yang lebih baik [10]. Sedangkan pada penelitian [8], dihasilkan sistem rekomendasi dengan metode *Content Based-Filtering* yang mencari kemiripan antara kata kunci dan beberapa atribut dalam sebuah film. Penelitian ini memanfaatkan pembobotan kata dengan *TF-IDF* dan menghitung kemiripan antara kata kunci dengan *genre*, sinopsis, dan judul dengan menggunakan *Cosine Similarity*. Oleh karena itu, peneliti mengajukan penelitian dengan judul Sistem Rekomendasi Drama Korea Berdasarkan Prediksi *Rating* dan Kemiripan Sinopsis Menggunakan *Singular Value Decomposition* dan *Term Frequency-Inverse Document Frequency*. Sehingga penelitian ini berfokus untuk menggunakan kemiripan sinopsis sebagai dasar dalam mencari rekomendasi drama korea, dengan prediksi *rating* yang digunakan sebagai referensi *user*.

II. METODE PENELITIAN

Penelitian ini memiliki dua tahapan, yaitu proses pembuatan model prediksi *rating* seperti pada Gbr 1 dan proses mendapatkan rekomendasi drama korea seperti pada Gbr 2.



Gbr. 1 Proses Pembuatan Model Prediksi *Rating*

Penjelasan proses pada Gbr. 1 adalah sebagai berikut.

A. Dataset

Peneliti telah mengambil dataset drama korea dari *website Kaggle* atau bisa diakses melalui link berikut https://www.kaggle.com/code/mabelhsu/ analyzing-past-kdrama-for-2024-s-big-hit/input. Pada *dataset* tersebut terdapat empat *file* yaitu *korean drama.csv*, *recommendations.csv*, *reviews.csv*, dan *wiki_actors.csv*. *File* yang akan digunakan hanya dua *file* saja, yaitu *korean_drama.csv* dan *reviews.csv*

B. Cleaning Data

Proses ini mengambil data yang diperlukan untuk penelitian, dan menghilangkan data kosong. Pada file *Korean_drama.cvs* akan digunakan tiga atribut atau kolom saja, yaitu *kdrama_id, drama_name,* dan *synopsis*. Kolom *drama_name* akan diubah menjadi *title,* lalu menghilangkan data yang kosong dan tersisa 1584 drama korea. Pada file *reviews.csv*, hanya tiga kolom yang akan digunakan, yaitu *user_id, title,* dan *overall_score*. Jumlah *user* yang memberikan *rating* sebanyak 4562 *user*. Namun, peneliti akan mengambil *user* yang paling sering memberikan *rating*, dengan minimal jumlah *rating* 7 ke atas, dan menyisakan 225 *user*. Kolom *overall_score* akan diubah menjadi *rating*. Karena peneliti memerlukan sinopsis dari drama korea untuk mencari kecocokan dengan deskripsi singkat yang dimasukkan *user*, maka perlu menggabungkan dua data di atas menjadi satu. Dari proses penggabungan data ini, karena tidak semua drama korea pernah ditonton oleh *user*, maka jumlah drama korea berkurang menjadi 1154 drama, dengan total *rating* sebanyak 4143 *rating*.

C. Preprocessing

1) *Case Folding: Case folding* dilakukan dengan merubah tiap kata pada sinopsis dan deskripsi singkat menjadi huruf kecil, menghilangkan angka, dan menghilangkan simbol.
2) *Tokenization, Lemmatization,* dan *Stopwords Removal:* Pada tahap ini dilakukan pemisahan kata, merubah ke bentuk dasarnya, dan menghilangkan *stopwords.*
3) *Penggabungan Kata:* Penggabungan kata dasar menjadi dalam satu dimensi *array* untuk tiap drama korea dilakukan agar memudahkan dalam melakukan tahap selanjutnya yaitu *TF-IDF*.

D. Pembagian Data

Dataset dibagi menjadi data *train* dan data *test*, dengan menerapkan lima skenario pembagian data, sebagai berikut :

1) 50% sebagai data *train* dan 50% sebagai data *test*
2) 60% sebagai data *train* dan 40% sebagai data *test*
3) 70% sebagai data *train* dan 30% sebagai data *test*.
4) 80% sebagai data *train* dan 20% sebagai data *test*.
5) 90% sebagai data *train* dan 10% sebagai data *test.*

E. Menerapkan Singular Value Decomposition

Pada penelitian ini menggunakan algoritma *Singular Value Decomposition* (*SVD*) yang merupakan salah satu teknik dari *Matrix Factorization. SVD* melakukan dekomposisi pada matriks A (matriks interaksi *user-item*) dengan dimensi $t\ x\ n$ menjadi perkalian tiga matriks U, S, dan V^T seperti pada Persamaan (1).

$$A = U \cdot S \cdot V^T \quad (1)$$

Dimana U dan V adalah matriks *orthogonal* dengan dimensi $t\ x\ r$ dan $r\ x\ n$. Sedangkan S atau $\Sigma\ (sigma)$ adalah matriks diagonal $r\ x\ r$ yang merupakan nilai singular dari matriks A dengan nilai diagonalnya positif dalam urutan menurun [11]. Pada penelitian ini, matriks U merupakan matriks user, dengan kolom-kolomnya merupakan basis *orthonormal* (tegak lurus) untuk kolom-kolom pada matriks A. Matriks S atau matriks $\Sigma\ (sigma)$ merupakan matriks diagonal berisikan nilai *singular* positif, yang merepresentasikan seberapa penting konsep tertentu pada matriks A. Nilai singular diurutkan dari yang terbesar ke terkecil. Sedangkan matriks V^T adalah matriks *item*, dengan baris-barisnya merupakan basis *orthonormal* (tegak lurus) untuk baris-baris pada matriks A.

Karena tidak semua user memberikan *rating* terhadap *item*, maka terdapat kemungkinan matriks memilki tingkat sparse yang tinggi (jumlah data kosong yang tinggi), namun pada *SVD* dapat mengatasi masalah pada matriks interaksi *user-item* dengan tingkat sparse yang tinggi [12]. Karena *SVD* menyediakan rumus prediksi *rating* seperti pada Persamaan (2)

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T \cdot p_u \quad (2)$$

Dimana $\hat{r}_{ui}$ adalah prediksi *rating*, μ adalah *global mean* atau rata-rata keseluruhan dari *rating*, b_u dan b_i masing-masing adalah bias *user* dan bias *item*, q_i^T merupakan *matrix item factor*, dan p_u adalah *matrix user factor*. Jika *user* u tidak diketahui, maka bias b_u dan faktornya p_u diasumsikan nol. Hal yang sama berlaku untuk *item* i dengan b_i dan q_i [3]. Penelitian ini memanfaatkan *library Surprise* dari *Python* untuk membuat model prediksi *rating* menggunakan algoritma *SVD.*

F. Evaluasi

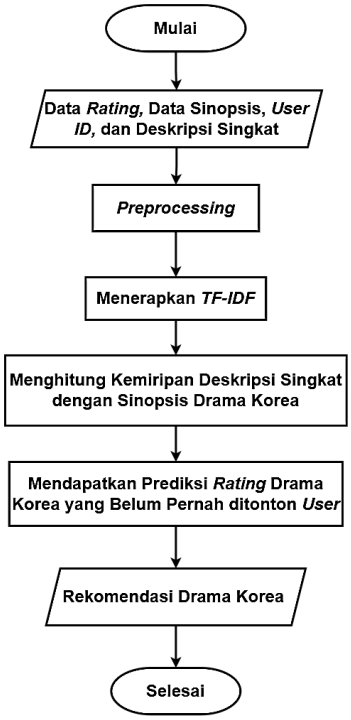
Pada penelitian ini menggunakan metrik evaluasi *Root Mean Square Error* (*RMSE*) untuk menghitung tingkat kesalahan prediksi. *RMSE* adalah metrik evaluasi yang banyak digunakan dan efektif untuk mengukur kinerja sistem rekomendasi, dapat dihitung menggunakan Persamaan (3)

$$RMSE = \sqrt{\frac{1}{|\hat{R}|}\sum_{\widehat{r_{ui}}\in\hat{R}}(r_{ui} - \widehat{r_{ui}})^2} \quad (3)$$

Dimana $\hat{R}$ adalah jumlah seluruh data yang diprediksi, r_{ui} adalah *rating* asli *user* u terhadap *item* i, dan $\widehat{r_{ui}}$ adalah prediksi *rating user* u terhadap *item* i. Hasil *RMSE* menunjukkan perbedaan antara hasil prediksi dan data aktual [11].

G. Model Prediksi

Setelah melalui tahap evaluasi, maka model terbaik telah berhasil dibuat. Model prediksi ini berisi prediksi *rating user* terhadap drama korea. Pada model ini bisa dilakukan pemanggilan nilai prediksi *rating* berdasarkan *user id* dan *kdrama id*.



Gbr. 2 Proses mendapatkan rekomendasi drama korea

Penjelasan proses pada Gbr. 2 adalah sebagai berikut.

A. Data Rating, Data Sinopsis, User ID, dan Deskripsi Singkat

Pada tahap untuk mencari rekomendasi dibutuhkan data *rating* dan data sinopsis. Data *rating* berisikan penilaian *user* terhadap drama korea, dengan nilai *rating* 1-10. Penggunaan nilai *rating* 1-10. Data sinopsis yang dipanggil atau di-*import* pada alur ini adalah data sinopsis yang sudah dilakukan *preprocessing* sebelumnya (Gbr. 1). Proses *preprocessing* pada data sinopsis dilakukan di awal penelitian karena hasil akhir prosesnya disimpan pada file csv. Sehingga pada tahap atau alur ini tidak perlu melakukan proses *preprocessing* lagi, dan bisa menjadi lebih efisien. *User id* yang dimasukkan *user* akan digunakan sebagai parameter untuk mencari drama korea yang belum pernah ditonton oleh *user,* dan mencari nilai prefiksi *rating* untuk *user id* tersebut. Sedangkan deskripsi singkat, digunakan sebagai parameter untuk mencari drama korea dengan sinopsis yang hampir mirip atau serupa dengan deskripsi singkat yang dimasukkan *user*.

B. Preprocessing

Deskripsi singkat yang telah dimasukkan *user* akan dilakukan *preprocessing*, sebelum nanti dilakukan perhitungan TF-IDF dan dicari kemiripannya dengan sinopsis drama korea.

Proses *preprocessing* yang dilakukan untuk deskripsi singkat sama dengan proses *preprocessing* pada data sinopsis drama korea, yaitu merubah deskripsi singkat menjadi format *lowercase* serta menghilangkan simbol dan angka. Kemudian melakukan *tokenization, lemmatization,* dan *stopwords removal.* Dan yang terakhir, menggabungkan kata menjadi satu dimensi array, supaya memudahkan dalam melakukan tahap selanjutnya yaitu TF-IDF. Proses *preprocessing* pada deskripsi singkat dilakukan di tahap ini, karena deskripsi singkat bisa berubah-ubah tergantung yang dimasukkan *user*. Sehingga setiap *user* memasukkan deskripsi singkat, perlu dilakukan proses *preprocessing*.

C. Menerapkan TF-IDF

Pada penelitian ini, penerapan *TF-IDF* memanfaatkan modul dari *sklearn* yaitu *TFIDFVectorizer*. Tujuan dari *TF-IDF* adalah untuk menghitung dan mengetahui bobot dari suatu kata pada data sinopsis dan deskripsi singkat. *TF* merupakan frekuensi atau jumlah kemunculan kata pada suatu dokumen. Sedangkan *IDF* merupakan tingkat kepentingan suatu kata. Kata yang sering muncul pada banyak dokumen dianggap tidak penting, sebaliknya jika kata tersebut langka (tidak banyak dijumpai pada banyak dokumen) maka kata tersebut memiliki kepentingan yang besar pada dokumen tersebut. *IDF* didapatkan dengan melakukan perhitungan logaritma dari jumlah keseluruhan dokumen dibagi dengan jumlah dokumen yang mengandung kata atau *term*, seperti pada Persamaan (4)

$$idf_t = \log_{10} \frac{N}{df_t} \quad (4)$$

Keterangan :
N : banyaknya seluruh dokumen
df_t : banyaknya dokumen yang mengandung *term* (t)

Perkalian *TF* dan *IDF* menghasilkan bobot *TF-IDF*, dapat dilihat pada Persamaan (5).

$$W_{t,d} = tf_{t,d} \,.\, idf_t \quad (5)$$

Keterangan :
$W_{t,d}$: bobot *term* (t) pada dokumen (d) (TF-IDF)
$tf_{t,d}$: jumlah kemunculan *term* (t) dalam dokumen (d)
idf_t : nilai *Inverse Document Frequency*

Hasil perhitungan bobot kata pada tiap data sinopsis berbeda, tergantung dari seberapa sering kata tersebut muncul pada suatu data sinopsis. Jika nilai bobot suatu kata mendekati satu, maka semakin bagus kata tersebut digunakan sebagai kata penyusun dalam deskripsi singkat yang dimasukkan *user*. Hasil perhitungan *TF-IDF* antara data sinopsis dan deskripsi singkat nanti akan dihitung kemiripannya dengan menggunakan *Cosine Similarity* seperti pada Persamaan (6).

$$sim(d_j, q) = \frac{d_j \cdot q}{\|d_j\| \|q\|} = \frac{\sum_{i=1}^{N} w_{i,j} w_{i,q}}{\sqrt{\sum_{i=1}^{N} w_{i,j}^2} \sqrt{\sum_{i=1}^{N} w_{i,q}^2}} \quad (6)$$

Keterangan :
$sim(d_j, q)$: similaritas dokumen ke-j dengan deskripsi singkat
d_j : dokumen ke-j
q : deskripsi singkat
$w_{i,j}$: bobot ke-i pada dokumen ke-j
$w_{i,q}$: bobot ke-i pada deskripsi singkat q

D. Menghitung Kemiripan Deskripsi Singkat dengan Sinopsis Drama Korea

Deskripsi singkat dan data sinopsis yang telah dilakukan preprocessing serta diberikan bobot pada tiap katanya dengan menggunakan TF-IDF. Kemudian dihitung kemiripan antar keduanya, menggunakan cosine similarity dengan rumus pada Persamaan (5).

E. Mendapatkan Prediksi Rating Drama Korea yang Belum Pernah ditonton User

Model prediksi yang telah dibuat sebelumnya telah menghitung prediksi *rating user* terhadap drama korea. Pada tahap ini peneliti akan mendapatkan prediksi *rating* dari drama korea yang belum pernah ditonton oleh *user* dengan memanfaatkan *user id* yang telah dimasukkan *user*. *Output* dari proses ini adalah kumpulan drama korea yang belum pernah ditonton oleh *user* dengan menampilkan atribut prediksi *rating.*

F. Rekomendasi Drama Korea Berdasarkan Prediksi Rating dan Kemiripan Sinopsis

Setelah dilakukan perhitungan kemiripan menggunakan *cosine similarity*, maka dihasilkan rekomendasi drama korea yang diurutkan berdasarkan tingkat kemiripan sinopsis. Serta menampilkan prediksi *rating* sebagai referensi *user*. Hasil rekomendasi yang diberikan bisa diurutkan berdasarkan tingkat kemiripan sinopsis atau berdasarkan prediksi *rating* tertinggi.

III. HASIL DAN PEMBAHASAN

A. Implementasi SVD dalam Memberikan Rekomendasi Drama Korea Berdasarkan Prediksi Rating

Pada penelitian ini, *library* utama yang digunakan adalah *library Pandas, Sklearn,* dan *Surprise* dari *Python*. Di dalam *Surprise* terdapat beberapa *module* yang digunakan yaitu *SVD*, *Reader, Dataset, accuracy RandomizedSearchCV,* dan *GridSearchCV*. *Reader* digunakan untuk membaca file , yang hanya berisikan nilai *rating,* spesifik sesuai dengan nilai *rating* yang ada pada dataset. *RandomizedSearchCV* dan *GridSearchCV*, digunakan untuk melakukan tuning *hyperparameter* dengan memberikan beberapa pilihan parameter yang nanti akan digunakan untuk melatih suatu model. Parameter yang terpilih merupakan kombinasi parameter terbaik yang diharapkan bisa memberikan hasil *training* yang lebih baik pula.

Data *rating* pada algoritma *SVD* akan dirubah menjadi tiga matriks dengan dimensi yang lebih kecil, yaitu matriks U, S, dan V^T. Gbr. 3 adalah matriks U atau matriks *user* berisikan informasi nilai laten faktor dari *user*, Gbr. 4 adalah matriks S atau matriks Σ ($sigma$) berisikan nilai *singular* positif, yang merepresentasikan seberapa penting konsep tertentu pada data *rating,* dan Gbr. 5 adalah matriks V^T atau matriks *item* berisikan informasi nilai laten faktor dari *item.* Ketiga matriks ini akan digunakan dalam perhitungan SVD untuk mencari *matrix item factor* dan *matrix user factor*, dan mencari prediksi *rating*.

	0	1	2	3	4	5	6	7	8	9	...
0	0.012285	-0.008662	0.004121	0.006894	-0.010091	0.008540	-0.019312	0.003253	-0.012815	-0.015078	...
1	0.004188	-0.002500	-0.003317	0.000373	-0.002471	0.006169	-0.024161	-0.001045	-0.000154	-0.000778	...
2	0.009907	0.016556	-0.019784	0.012145	-0.080316	0.024690	-0.001612	-0.002926	0.000382	-0.026449	...
3	0.041173	0.025237	-0.017118	-0.002195	0.029748	-0.024043	-0.033229	-0.009484	0.003766	-0.011741	...
4	0.095640	-0.011765	0.254357	0.094779	-0.104866	-0.045344	-0.007936	0.053947	-0.051578	0.593891	...
...	...	...	...	...	...	...	...	...	...	...	...

Gbr. 3 Matriks U atau matriks *user*

	0	1	2	3	4	5	6	7	8	9	...
0	123.473859	0.00000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	...
1	0.000000	98.91449	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	...
2	0.000000	0.00000	89.958744	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	...
3	0.000000	0.00000	0.000000	80.972603	0.000000	0.0	0.0	0.0	0.0	0.0	...
4	0.000000	0.00000	0.000000	0.000000	77.879953	0.0	0.0	0.0	0.0	0.0	...
...	...	...	...	...	...	...	...	...	...	...	...

Gbr. 4 Matriks S atau matriks Σ ($sigma$)

	0	1	2	3	4	5	6	7	8	9	...
0	0.025927	-0.001477	0.089914	0.041776	-0.018937	0.003823	-0.001455	0.006835	-0.049828	0.051762	...
1	0.034238	0.050570	0.008698	-0.073708	-0.030357	0.048705	-0.095370	-0.012761	0.008619	0.017176	...
2	0.050470	0.011655	-0.023841	0.025636	0.062443	-0.016560	0.027831	0.044846	-0.010148	0.024546	...
3	0.033307	-0.042015	-0.014451	-0.006898	0.009659	0.017452	0.006149	0.000772	0.000262	0.007283	...
4	0.039797	-0.048069	-0.010927	-0.007894	0.009975	0.020028	0.008597	0.004050	0.003332	0.011231	...
...	...	...	...	...	...	...	...	...	...	...	...

Gbr. 5 Matriks V^T adalah matriks item

Pembagian data dilakukan secara acak menggunakan *module train_test_split*, dengan lima skenario pembagian data.

1) 50% data *train* = 2071 data, 50% data *test* = 2072 data
2) 60% data *train* = 2485 data, 40% data *test* = 1658 data
3) 70% data *train* = 2900 data, 30% data *test* = 1243 data
4) 80% data *train* = 3314 data, 20% data *test* = 829 data
5) 90% data *train* = 3728 data, 10% data *test* = 415 data

Seluruh data *rating* akan dilakukan proses *tuning* dengan menggunakan dua algoritma pencarian, yaitu *RandomizedSearchCV* dan *GridSearchCV.* Proses *tuning* merupakan proses untuk mencari dan menemukan kombinasi parameter yang paling optimal atau disebut sebagai *hyperparameter* [13]. Dapat diketahui pada Tabel I, hasil pencarian parameter terbaik atau *hyperparameter* menggunakan dua algoritma pencarian *RandimizedSearchCV* dan *GridSearchCV*. *Hyperparameter* dengan nilai *tuning* RMSE paling kecil akan digunakan dalam proses *training* dan *testing* model prediksi. Pada penelitian ini *hyperparameter* dengan nilai *tuning RMSE* terkecil ditemukan dengan menggunakan algoritma pencarian *GridSearchCV*.

TABEL I
HASIL PROSES TUNING

Algoritma Pencarian	*Hyperparameter*	*Tuning* RMSE
RandomizedSearchCV	*Factors* (n_factors) = 50 *Epochs* (n_epochs) = 30 *Learning* rate (lr_all) = 0.01 *Regularization* (reg_all) = 0.4	1.561353
GridSearchCV	*Factors* (n_factors) = 50 *Epochs* (n_epochs) = 30 *Learning* rate (lr_all) = 0.01 *Regularization* (reg_all) = 0.2	1.543592

Setelah parameter terbaik ditemukan, proses selanjutnya adalah *training* dan *testing* pada model SVD. *Training* dan *testing* dilakukan pada kelima skenario yang telah ditentukan sebelumnya. Kemudian akan dicari dan dibandingkan nilai *RMSE*-nya, skenario dengan nilai *RMSE* terkecil adalah skenario yang akan digunakan untuk model prediksi. Hasil nilai RMSE dirangkum pada Tabel II, dengan nilai *test* RMSE terkecil dihasilkan pada skenario kelima (90:10) yaitu 1.493423. Semakin kecil nilai RMSE maka semakin baik pula model prediksi tersebut [14]. Sehingga pada penelitian ini akan menggunakan model yang dibangun dengan perbandingan pembagian data 90:10, dengan menggunakan kombinasi *hyperparameter* yaitu *n_factors* = 50, *n_epochs* = 30, *lr_all* = 0.01, dan *reg_all* = 0.2. Kemudian model ini disimpan dalam bentuk *file*, yang nanti akan digunakan untuk mencari prediksi *rating user*. Proses menyimpan model menggunakan *module* yang disediakan *Python* yaitu, *pickle*.

TABEL II
PERBANDINGAN NILAI RMSE

Skenario	*Train* RMSE	*Test* RMSE
50:50	0.754458	1.608769
60:40	0.767468	1.58099
70:30	0.768886	1.615144
80:20	0.774086	1.621285
90:10	0.780879	1.493423

B. Implementasi TF-IDF dalam Memberikan Rekomendasi Drama Korea Berdasarkan Kemiripan Sinopsis

Data sinopsis dan deskripsi singkat yang telah melalui tahap *prepocessing* di-*import* dan dilakukan perhitungan TF-IDF pada data tersebut. Peneliti menggunakan modul dari *sklearn* yaitu *CountVectorizer* dan *TFIDFVectorizer*. Modul *CountVectorizer* digunakan untuk mencari nilai *Term Frequency* (TF). *TFIDFVectorizer* digunakan untuk mencari nilai *Inverse Document Frequency* (IDF). Dapat dilihat dari Gbr. 6 adalah nilai TF atau frekuensi kemunculan kata pada suatu dokumen dan nilai IDF atau kepentingan suatu kata. Kemudian kedua nilai ini akan digunakan untuk mencari nilai *Term Frequency - Inverse Document Frequency* (TF-IDF).

	0	1	2	3	4	5	6	7	8	9	...
abandon	0	0	0	0	0	0	0	2	0	1	...
ability	0	0	1	0	0	0	0	0	0	0	...
able	0	0	1	0	0	1	0	0	0	0	...
accept	0	0	0	0	0	1	0	0	0	0	...
accident	0	0	0	0	0	0	0	0	1	0	...
...	...	...	...	...	...	...	...	...	...	...	...

	IDF
abandon	4.833845
ability	3.893838
able	3.776055
accept	4.756884
accident	3.818615
...	...

Gbr. 6 Nilai *TF* dan nilai *IDF*

Hasil TF-IDF pada data sinopsis seperti pada Gbr. 7, pada hasil pembobotan kata, angka yang berada dalam tanda kurung merepresentasikan *index* dari data sinopsis dan *index* dari kata, sedangkan angka disebelah kanan merepresentasikan bobot kata pada suatu data sinopsis. Diperoleh jumlah kata unik pada data sinopsis yaitu sebanyak 1004 kata unik. Dari data sinopsis, didapatkan jumlah bobot tiap kata pada tiap kalimat sinopsis sebanyak 43267 bobot kata. Dimana bobot tiap kata berbeda dengan bobot kata lainnya.

```
Jumlah kata unik : 1004 kata

Daftar Kata Unik yang Muncul :
['abandon' 'ability' 'able' ... 'youth' 'youtube' 'yu']

Jumlah bobot tiap kata pada tiap kalimat : 43267 bobot kata

Hasil Pembobotan Tiap Kata pada Tiap Kalimat :
<Compressed Sparse Row sparse matrix of dtype 'float64'
        with 43267 stored elements and shape (1155, 1004)>
  Coords        Values
  (0, 30)       4.618733844747176
  (0, 39)       3.925586664187231
  (0, 72)       4.29152093353876
  (0, 106)      3.8042258071829638
  (0, 109)      4.651523667570167
  (0, 134)      4.181520038324432
  (0, 151)      5.0081986115089
  (0, 155)      5.056988775678332
  (0, 164)      2.882237054194171
  (0, 193)      5.413663719617064
  (0, 202)      4.756884183227994
  (0, 207)      4.874667218884377
  (0, 249)      5.108282070065882
  (0, 334)      2.9901260162053562
```

Gbr. 7 Eksplorasi Hasil *TF-IDF*

Sebelum melakukan perhitungan *TF-IDF* dan perhitungan *cosine similarity*, deskripsi singkat terlebih dahulu dilakukan *preprocessing*. Proses *preprocessing* pada deskripsi singkat dilakukan setiap *user* memasukkan deskripsi singkat untuk mencari rekomendasi. Peneliti membuat lima skenario percobaan dengan memasukkan lima deskripsi singkat yang berbeda seperti pada Tabel III. Pada skenario pertama, deskripsi singkat membahas mengenai drama korea yang bertema kehidupan kedokteran, medis, dan operasi. Skenario kedua adalah drama korea tentang hubungan pertemanan dan percintaan pada masa remaja dikehidupan sekolah. Drama korea yang berlatar kehidupan jaman Dinasti Joseon (kerajaan Korea), akan digunakan pada skenario ketiga. Pada skenario keempat membahas mengenai drama korea yang menceritakan masalah kehidupan dari pemainnya di kehidupan yang nyata. Pada skenario kelima, adalah drama dengan tema masalah hukum dan kepolisian.

TABLE III
DESKRIPSI SINGKAT

No	Deskripsi Singkat
1.	*Medical, surgical procedures, and personal struggles of doctors.*
2.	*Romance and friendship in high school or love of teenage.*
3.	*Life in the Joseon Dynasty, political power, and historical of king Korea.*
4.	*Understanding life's purpose, around life challenges, and the search for true meaning of life.*
5.	*Law, police, cases, and investigation.*

TABLE IV
HASIL PERCOBAAN DESKRIPSI SINGKAT

No Des. Singkat	Judul Drama	Sinopsis	Kemiripan Sinopsis
1	*Doctor John*	*""Doctor John" is a medical drama about doctors specializing in pain management. In a refreshing take on the genre, "Doctor John" will portray the doctors' search for the cause of their patients' mysterious pain as a thrilling chase, almost like a detective hunting down the perpetrator behind an unsolved crime. Cha Yo Han is a genius anesthesiologist, who is also the youngest professor at his medical school. The brilliant doctor goes by the nickname "10 Seconds," a reference to his ability to diagnose his patients in the 10 seconds that it takes for them to enter an examination room and walk to their seat.Kang Shi Young is a legendary anesthesiologist, who was always at the top of her class throughout medical school. The talented doctor inherited her skills from her cool-headed, rational mother, while she inherited her empathy, listening skills, and warm bedside manner from her father."*	0.456945109 (45%)
2	*A-Teen*	*"People say that 18 is a carefree age, but all the moments were too serious to say I was not worried. This is a short teenage school romance web drama."*	0.346867091 (34%)
3	*Six Flying Dragons*	*"A fictional, historical drama about the ambition, rise and fall of real and fictional characters based around Yi Bang Won. Yi Bang Won was the third king of the Joseon Dynasty in Korea and the father of King Se Jong the Great. He helped his father, King Tae Jo, establish the Joseon Dynasty."*	0.534332776 (53%)
4	*All is Well*	*"A drama about people who face hate and conflict while learning to forgive and understand the meaning of love and happiness."*	0.300057749 (30%)
5	*Hidden Identity*	*"The story of the investigative team Number 5 specially organized to actively combat crime through stake outs, wiretapping, communications monitoring, and deep-cover infiltration. A secret police unit that bends the limits of the law in order to fight crime, the new department comes in addition to the existing departments for criminal investigations, special investigations, criminal affairs, and white collar crime."*	0.358191663 (35%)

Dapat dilihat dari hasil percobaan pada Tabel IV, nilai kemiripan tertinggi dihasilkan pada percobaan ke tiga dengan kemiripan sebesar 53%. Hal ini disebabkan karena sebanyak lima kata pada deskripsi singkat ditemukan pada data sinopsis drama korea dengan judul '*Six Flying Dragons*', yaitu *historical, king, joseon, dynasty*, dan korea. Dimana kata joseon, *dynasty*, dan, *king* memiliki nilai IDF yang tinggi (dianggap sebagai kata yang penting), berturut-turut yaitu 4.315051431, 5.10828207, 4.441803137. Ketiga kata ini juga muncul sebanyak 2 dan 3 kali pada sinopsis drama ini, sehingga perhitungan *cosine similarity*-nya tinggi. Sedangkan hasil nilai kemiripan terendah adalah pada percobaan ke empat, yaitu sebesar 30%. Disebabkan dari keseluruhan kata pada deskripsi singkat hanya 2 kata yang muncul pada data sinopsis yaitu *understand* dan *meaning*. Dan dua kata tersebut masing-masing memiliki nilai IDF sebesar 4.874667219 dan 4.224079653.

Hasil kemiripan menggunakan *TF-IDF* dan *cosine similarity* bergantung berdasarkan kemiripan kata dasar yang sama persis, bukan berdasarkan kemiripan makna kata. Sehingga, meskpun terdapat kata yang memiliki makna yang sama namun kata tersebut berbeda bentuk kata dasarnya, maka tetap dikatakan sebagai kata yang berbeda. Hasil rekomendasi dikatakan cukup rendah karena hanya beberapa kata pada deskripsi singkat yang muncul dalam data sinopsis. Hal ini juga dipengaruhi oleh panjang dokumen yang dibandingkan berbeda-beda dan data sinopsis yang bervariasi, sehingga dapat mempengaruhi jumlah kata yang muncul dan tingkat kepentingan kata itu sendiri.

C. Tingkat Akurasi Prediksi *Rating* dari Drama Korea yang Direkomendasikan kepada Pengguna

Tingkat akurasi prediksi *rating* dipengaruhi dengan nilai RMSE. Semakin kecil nilai RMSE, maka semakin kecil juga nilai selisih antara *rating* asli dengan prediksi *rating*. Jika nilai selisih kecil, maka nilai prediksi *rating* hampir mendekati nilai *rating* aslinya. Dalam mencari rekomendasi berdasarkan prediksi *rating*, parameter yang diperlukan adalah *user_id*. Pada penelitian ini, menggunakan percobaan *user_id* = 86, 64, dan 207. Pada Gbr 8 dapat dilihat hasil prediksi *rating* pada drama korea yang telah di tonton oleh *user_id* = 86, 64, dan 207, dengan ditampilkan nilai selisih antara prediksi *rating* dan *rating* aslinya. Nilai selisih terkecil adalah 0.392477, 0.000977, dan 0.004264, dan nilai selisih terbesarnya adalah 1.141524, 0.911507, dan 2.204109.

	user_id	kdrama_id	real_rating	pred_rating	selisih
1	86	810	8.5	8.107523	0.392477
2	86	745	5.5	6.030512	0.530512
3	86	866	8.0	7.366407	0.633593
4	86	583	9.0	8.344656	0.655344
5	86	429	9.0	8.140129	0.859871
6	86	428	5.5	6.546071	1.046071
7	86	1113	8.5	7.417185	1.082815
8	86	1070	3.5	4.641524	1.141524

	user_id	kdrama_id	real_rating	pred_rating	selisih
1	64	1120	9.0	9.000977	0.000977
2	64	103	10.0	9.849272	0.150728
3	64	1094	10.0	9.849240	0.150760
4	64	897	9.0	9.238376	0.238376
5	64	514	10.0	9.639347	0.360653
6	64	213	10.0	9.596047	0.403953
7	64	1077	10.0	9.574215	0.425785
8	64	991	10.0	9.566735	0.433265
9	64	144	10.0	9.379230	0.620770
10	64	417	10.0	9.088493	0.911507

	user_id	kdrama_id	real_rating	pred_rating	selisih
1	207	291	9.0	9.004264	0.004264
2	207	938	8.0	7.861017	0.138983
3	207	1073	9.5	9.294999	0.205001
4	207	249	9.0	8.771587	0.228413
5	207	516	9.0	8.765086	0.234914
6	207	264	9.0	8.714338	0.285662
7	207	307	9.5	8.898899	0.601101
8	207	658	3.0	3.731148	0.731148
9	207	35	9.5	8.704287	0.795713
10	207	116	10.0	9.203354	0.796646
11	207	954	9.5	8.553122	0.946878
12	207	826	9.0	8.006216	0.993784
13	207	809	3.0	5.204109	2.204109

Gbr. 8 Prediksi *rating* pada drama korea dengan *user id* = 86, 64, dan 207

Sistem rekomendasi yang dibangun pada penelitian ini adalah perpaduan dari metode *content based filtering* dan metode *collaborative filtering*, yang biasa disebut dengan *hybrid filtering*. Jenis *hybrid filtering* yang digunakan yaitu *switching hybrid filtering*, dimana metode ini akan menggabungkan hasil rekomendasi drama dari kedua metode yang peneliti gunakan. *User* akan mendapatkan dua hasil rekomendasi berdasarkan urutan kemiripan sinopsis dan prediksi *rating*. Penggunaan metode *hybrid filtering* digunakan untuk membuat sistem rekomendasi dengan hasil yang lebih baik dari menggabungkan dua metode atau lebih. Tabel V merupakan hasil rekomendasi drama korea berdasarkan prediksi *rating*, yang dimana rekomendasi ini adalah hasil dari rekomendasi sebelumnya yaitu rekomendasi drama korea berdasarkan kemiripan sinopsis. Seperti yang sudah dijelaskan pada alur untuk mendapatkan rekomendasi drama korea berdasarkan prediksi *rating* sebelumnya, *user* diperintahkan untuk memasukkan *user_id*, yang kemudian digunakan untuk mencari prediksi *rating* menggunakan model prediksi yang telah dibuat sebelumnya. Sebagai percobaan, peneliti memilih *user_id* = 86 yang ingin dicari prediksi *rating*-nya.

Hasil rekomendasi ditampilkan dengan beberapa atribut yaitu, *kdrama id*, *title*, *synopsis*, *similarity sinopsis*, *prediction rating*. *Similarity_sinopsis* adalah tingkat kemiripan sinopsis dengan deskripsi singkat yang sebelumnya dimasukkan oleh user. *Prediction rating* adalah nilai prediksi *rating user* terhadap suatu drama. Hasil rekomendasi drama korea berdasarkan kemiripan sinopsis akan berubah urutannya, bila rekomendasi drama korea diurutkan berdasarkan prediksi *rating*. Karena prediksi *rating* dihitung berdasarkan riwayat dan preferensi suatu *user* dalam memilih atau menilai suatu drama korea. Pada Tabel V adalah hasil rekomendasi drama korea untuk *user id* = 86, dimana hasil rekomendasi ini masih berhubungan dengan hasil rekomendasi sebelumnya, yaitu berdasarkan kemiripan sinopsis. Terdapat lima hasil rekomendasi untuk *user id* = 86, berdasarkan lima percobaan deskripsi singkat.

TABLE V
HASIL REKOMENDASI DRAMA KOREA BERDASARKAN PREDIKSI *RATING* DAN KEMIRIPAN SINOPSIS, UNTUK *USER ID* = 86

no	*kdrama id*	*title*	*synopsis*	*similarity sinopsis*	*prediction rating*
1	689	*Doctor John*	*""Doctor John" is a medical drama about doctors specializing in pain management. In a refreshing take on the genre, "Doctor John" will portray the doctors' search for the cause of their patients' mysterious pain as a thrilling chase, almost like a detective hunting down the perpetrator behind an unsolved crime. Cha Yo Han is a genius anesthesiologist, who is also the youngest professor at his medical school. The brilliant doctor goes by the nickname "10 Seconds," a reference to his ability to diagnose his patients in the 10 seconds that it takes for them to enter an examination room and walk to their seat.Kang Shi Young is a legendary anesthesiologist, who was always at the top of her class throughout medical school. The talented doctor inherited her skills from her cool-headed, rational mother, while she inherited her empathy, listening skills, and warm bedside manner from her father."*	0.456945109	8.372232571
2	1001	*The Mermaid Prince: The Beginning*	*""The Mermaid Prince: The Beginning" will take viewers back in time to Woo Hyuk's high school days in order to uncover a secret.With its high school setting, the drama will also portray an entirely different kind of*	0.270833194	8.208728062

			romance from that of the original series. "The Mermaid Prince: The Beginning" will tell the love story that ensues when Jo Ara, a high school student who decides to give up on dating because she doesn't believe in love, ends up meeting Woo Hyuk.Yoon Gun is a guy that looks cold on the outside but is actually a romantic at heart and only has eyes for one person."		
3	291	*Joseon Exorcist*	*"During the latter part of the 14th century, King Taejong abolished the Goryeo Dynasty in order to establish the Kingdom of Joseon. Some were aware of the deal he made with an evil spirit — which in turn used their power to possess humans who would control and terrorize the world. The king's sons, Princes Chung Nyung and Yang Nyeong, are his main opposition, each with a different reason for standing in his way.(Source: DramaWiki)~~ The series was cancelled after two episodes."*	0.371293513	8.45894585
4	278	*Familiar Wife*	*"Cha Joo Hyuk works at a bank and has been married to Seo Woo Jin for five years. When a strange incident happens one day, Joo Hyuk makes a decision that impacts his life and those around him in unexpected ways. Suddenly, the life he had with Woo Jin and his best friend, Yoon Joong Hoo, is gone, and he is leading a very different life.How will his first love, Lee Hee Won, factor into his new life? And will it be possible to get his old life back?(Source: Viki)"*	0.250850418	7.774776357
5	302	*Beyond Evil*	*"Meet the two fearless men willing to go to extreme lengths in their pursuit of a serial killer that has shaken up their quiet city: Lee Dong Sik, a once capable detective, is now demoted to perform menial tasks at the Manyang Police Substation. Just as he is beginning to settle into a peaceful life, his new superior, Detective Han Joo Won, is transferred in. Joo Won is an elite detective whose father is the star candidate for the National Police Agency's next chief. Despite his admirable qualities and high social standing, Joo Won harbors a long-held secret.When a string of gruesome killings occurs, a pattern takes shape that is oddly reminiscent of some cold serial-murder cases from 20 years prior. Dong Sik and Joo Won team up, vowing to stop this vicious culprit in his tracks. As their investigation proceeds, they are forced to look far deeper than what the evidence suggests, questioning the culpability of all those surrounding the case, including their own.(Source: Kdramapal)"*	0.247556971	8.300642097

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan pada penelitian yang telah dilakukan, dapat diperoleh kesimpulan sebagai berikut:

1. Pada penelitian ini membuat model prediksi *rating* menggunakan *Singular Value Decomposition (SVD)* dengan lima skenario pembagian data yaitu 50:50, 60:40, 70:30, 80:20, dan 90:10. Model dilatih dengan kombinasi *hyperparameter n_factors* = 50, *n_epochs* = 30, *lr_all* = 0.01, dan *reg_all* = 0.2. Menghasilkan model prediksi terbaik pada skenario pembagian data 90:10, dengan nilai RMSE 1.493423.
2. Penggunaan metode *Term Frequency – Inverse Document Frequency* (*TF-IDF*) digunakan untuk memberikan bobot tiap kata pada data sinopsis dan deskripsi singkat, yang kemudian dihitung kemiripannya menggunakan *cosine similarity*. Peneliti menggunakan lima skenario deskripsi singkat yang dimasukkan untuk mencari drama korea dengan data sinopsis yang mirip. Hasil kemiripan tertinggi yang diperolah sebesar 53% pada percobaan deskripsi singkat ketiga dengan judul '*Six Flying Dragons*'.
3. Pada penelitian ini akurasi atau nilai selisih antara *rating* asli dan prediksi *rating* dengan percobaan *user id* = 86, 64, dan 207 berturut-turut adalah 0.392477, 0.000977, dan 0.004264. Hasil nilai selisih terbaik dari keseluruhan data hampir mendekati nol (0). Sehingga dapat disimpulkan bahwa hampir tidak ada kesalahan dalam mencari nilai prediksi *rating user* untuk *item* yang relevan.

V. SARAN

Penelitian ini masih perlu dan dapat dikembangkan kembali, supaya dapat memberikan hasil rekomendasi yang lebih baik lagi. Adapun saran peneliti untuk penelitian mendatang adalah sebagai berikut:

1. Eksplorasi dan evaluasi lebih dalam tentang algoritma *Singular Value Decompotion* dan *Term Frequency Inverse Document Frequency.*
2. Menggunakan algoritma lain untuk mencari prediksi *rating* dan kemiripan sinopsis sebagai perbandingan, dan supaya bisa memberikan rekomendasi yang lebih baik.
3. Menggunakan algoritma pendukung lainnya, seperti *Abstractive Text Summarization*, agar pencarian rekomendasi bisa berdasarkan inti atau kesimpulannya.

UCAPAN TERIMA KASIH

Ucapan terimakasih syukur alhamdulillah kepada Allah SWT atas rahmat dan keberkahannya selalu, kedua orang tua, dosen pembimbing, kakak, teman-teman, seluruh pihak yang ikut serta dalam penelitian ini, dan kepada peneliti sendiri yang sudah bertahan dengan baik untuk bisa menyelesaikan penelitian ini.

REFERENSI

[1] B. A. P. Yuliawan and G. E. Subakti, "Pengaruh Fenomena Korean Wave (K-Pop Dan K-Drama) Terhadap Perilaku Konsumtif Penggemarnya Perspektif Islam," vol. 18, no. 01, pp. 1829–6491, 2022, doi: 10.20414/jpk.v18i1.5195.

[2] R. Lumbantoruan, P. Simanjuntak, I. Aritonang, and E. Simaremare, "Topc-Camf: Sistem Perekomendasi Matrix Factorization Berbasis Top Context," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* |, vol. 11, no. 4, 2022.

[3] J. E. Prayogo, A. Suharso, and A. Rizal, "Analisis Perbandingan Model Matrix Factorization Dan K-Nearest Neighbor Dalam Mesin Rekomendasi Collaborative Berbasis Prediksi *Rating*," *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, p. 506, Dec. 2021, doi: 10.32493/informatika.v5i4.7379.

[4] Q. M. Rohima, D. N. Aini, and R. Setyarahajoe, "The Existence of Korean Music Industry (K-pop) on Changes in Communication Style and Consumptive Behavior of K-Popers in Surabaya Eksistensi Industri Musik Korea (K-Pop) pada Perubahan Gaya Berkomunikasi dan Perilaku Konsumtif K-Popers di Surabaya," *Kinesik*, vol. 11, no. 2, pp. 225–241, 2024, doi: 10.22487/ejk.v11i2.1375.

[5] R. Rosyidah Ilmi, F. Kurniawan, and S. Harini, "Prediksi *Rating* Film IMDb Menggunakan Decision Tree," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 4, pp. 791–798, 2023, doi: 10.25126/jtiik.2023106615.

[6] X. Li, H. Zhao, Z. Wang, and Z. Yu, "Research on Movie *Rating* Prediction Algorithms," in *2020 5th IEEE International Conference on Big Data Analytics, ICBDA 2020*, Institute of Electrical and Electronics Engineers Inc., May 2020, pp. 121–125. doi: 10.1109/ICBDA49040.2020.9101282.

[7] A. N. Laili, P. P. Adikara, and S. Adinugroho, "Rekomendasi Film Berdasarkan Sinopsis Menggunakan Metode Word2vec," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 6, pp. 6035–6043, 2019, [Online]. Available: http://j-ptiik.ub.ac.id

[8] M. Fajriansyah, P. P. Adikara, and A. W. Widodo, "Sistem Rekomendasi Film Menggunakan Content Based Filtering," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 6, pp. 2188–2199, 2021, [Online]. Available: http://j-ptiik.ub.ac.id

[9] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity Dan Pembobotan Tf-Idf System Penerimaan Mahasiswa Baru Pada Kampus Swasta," *Jurnal Bumigora Information Technology (BITe)*, vol. 3, no. 1, pp. 19–27, Jul. 2021, doi: 10.30812/bite.v3i1.1110.

[10] B. Teguh Imani and E. B. Setiawan, "Pengaruh Fenomena Korean Wave (K-Pop Dan K-Drama) Terhadap Perilaku Konsumtif Penggemarnya

Perspektif Islamrecommender System Based on Matrixs Factorization on Twitter Using Random Forest (Case Study: Movies on Netflix)," *Intl. Journal on ICT*, vol. 8, no. 2, pp. 11–21, 2022, doi: 10.21108/ijoict.v8i2.

[11] W. Shalannanda, R. F. Mulia, A. I. Muttaqien, N. R. Hibatullah, and A. Firdaus, "Singular Value Decomposition Model Application for E-commerce Recommendation System," *JITEL (Jurnal Ilmiah Telekomunikasi, Elektronika, dan Listrik Tenaga)*, vol. 2, no. 2, pp. 103–110, Sep. 2022, doi: 10.35313/jitel.v2.i2.2022.103-110.

[12] R. Siringoringo, Jamaluddin, and G. Lumbantoruan, "Sistem Perekomendasi Dengan Singular Value Decomposition Dan Teknik Similaritas Pearson Correlation," *Jurnal METHODIKA*, vol. 7, no. 1, pp. 19–24, 2021.

[13] J. Agung Nurcahyo and T. Bayu Sasongko, "Hyperparameter Tuning Algoritma Supervised Learning untuk Klasifikasi Keluarga Penerima Bantuan Pangan Beras," *Indonesian Journal of Computer Science*, vol. 12, no. 3, pp. 1351–1365, Jun. 2023.

[14] C. Wibisono, L. S. Haryadi, J. E. Widyaya, and S. L. Liliawati, "Sistem Rekomendasi Suku Cadang Berdasarkan Item Based Filtering," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, pp. 10–19, Apr. 2021, doi: 10.28932/jutisi.v7i1.3036.