

Anime Rekomendasi Menggunakan Collaborative Filtering

1st Iklil jayaperwira
Falkutas Informatika
Universitas Telkom
Bandung, Indonesia

ikliljayaperwira@student.telkomu
niversity.ac.id

2nd Agung Toto Wibowo
Falkutas Informatika
Universitas Telkom
Bandung, Indonesia

agungtotowibowo@telkomunivers
ity.ac.id

3rd Dade Nurjanah
Falkutas Informatika
Universitas Telkom
Bandung, Indonesia

dadenurjanah@telkomuniversity.a
c.id

Abstrak—Di era digital ini orang-orang semakin mudah mendapatkan hiburan yang mereka perlukan salah satunya adalah anime[1]. Anime merupakan animasi khas dari jepang anime bisa di buat baik di gambar menggunakan tangan atau menggunakan komputer. Anime menjadi salah satu hiburan yang banyak di sukai orang-orang di dunia, hal ini bisa di lihat dari Netflix salah satu layanan streaming yang besar mulai memasukkan anime ke dalam aplikasi dan situs mereka. Pada tahun 2021 sekarang terdapat kurang lebih 18350 anime baik yang sudah selesai maupun yang masih berlanjut[2]. hal ini membuat orang-orang yang sudah menyukai anime ataupun orang-orang yang baru ingin menonton anime kebingungan mencari anime yang sesuai dengan selera mereka karena itulah kita memerlukan sistem rekomendasi. Sistem rekomendasi merupakan sistem yang dibuat untuk membantu pengguna mendapatkan rekomendasi sebuah barang/informasi yang pengguna sukai/butuh dari banyaknya barang ataupun informasi yang ada. Rekomendasi yang di berikan di harapkan bisa memberikan bantuan pada pengguna untuk dapat menentukan pilihan yang akan di ambil. Dalam sistem rekomendasi sendiri terdapat banyak metode yang bisa di gunakan salah satunya adalah metode collaborative filtering yang di gunakan untuk mencari kesamaan item/ barang yang di carik oleh user lain[3] dengan algoritma yang digunakan adalah KNNWithMeans yang merupakan salah satu basic algoritma collaborative filtering[4], [5]. Pada penelitian ini dilakukan tiga skenario pengujian yang berguna untuk mendapatkan hasil rekomendasi terbaik dengan melakukan pengukuran MAE dan NDCG. Dapat di simpulkan metode collaboratif filtering dengan menggunakan algoritma KNNWithMeans mendapatkan rekomendasi yang cukup akurat dengan hasil MAE terbaik sebesar 0.8989 dan NDCG sebesar 0.2028.

Kata kunci—sistem rekomendasi, collaborative filtering

Abstract—In this digital era, it is easier for people to get the entertainment they need, one of which is anime[1]. Anime is a typical animation from Japan, anime can be drawn either by hand or using a computer. Anime is one of the entertainments that many people like in the world, this can be seen from Netflix, one of the big streaming services that has started to include anime into their applications and websites. In 2021 there are now approximately 18350 anime both completed and still ongoing[2]. this makes people who already like anime or

people who just want to watch anime confused looking for an anime that suits their taste that's why we need a recommendation system. A recommendation system is a system created to help users get recommendations for an item/information that users like/need from the number of items or information available. The recommendations given are expected to provide assistance to users to be able to determine the choices that will be taken. In the recommendation system itself, there are many methods that can be used, one of which is the collaborative filtering method which is used to find the similarity of items / items sought by other users[3]. The algorithm used is KNNWithMeans which is one of the basic collaborative filtering algorithms[4], [5]. In this study, three useful test scenarios were carried out to obtain the best recommendation results by measuring MAE and NDCG. It can be concluded that the collaborative filtering method using the KNNWithMeans algorithm obtained a fairly accurate recommendation with the best MAE result of 0.8989 and NDCG of 0.2028.

Keywords— recommendation system, collaborative filtering

I. PENDAHULUAN

Diera perkembangan digital ini orang-orang semakin mudah mendapatkan informasi baik yang berguna untuk menambah wawasan dan juga untuk hiburan salah satunya adalah anime[1]. Anime merupakan sebutan animasi atau gambaran yang dibuat di jepang anime biasanya memiliki gaya penggambaran yang unik dari pada animasi dari negara lain, kebanyakan anime memiliki ciri-ciri mata yang cukup besar, mulut yang lebih kecil dan hidung yang lebih mancung dari pada penggambaran dari animasi negara lain[6]. Anime bisa berupa sebuah komik atau gambaran dan juga bisa berupa vidio-vidio yang biasanya memiliki durasi 24menit per episode, anime makin dikenal di luar jepang hal ini bisa dilihat dari mulai banyaknya judul anime yang bermunculan di situs- situs terkenal seperti netflix, amazon prime, dan google-play hal ini juga bisa dilihat dari banyaknya acara pameran anime seperti anime expo, anime matsuri, comic fest ,dan lainnya.

Pada tahun 2021 terdapat kurang lebih 18350 anime yang sudah ada baik yang masih berlangsung maupun sudah selesai[2]. Anime memiliki banyak

genre yang dimiliki terdapat juga genre khusus yang dimiliki seperti shojo yang di mana anime yang dibuat menargetkan penontonnya anak perempuan terdapat juga genre shonen yang menargetkan penonton anak laki-laki. Dengan banyaknya genre dan banyaknya anime yang telah ada membuat orang-orang baik yang sudah sering menonton anime maupun yang baru ingin menonton kesusahan menentukan pilihan mereka, dikarenakan itulah diperlukan sistem yang bisa memberi rekomendasi anime ke user.

Sistem rekomendasi adalah sebuah algoritma atau program yang dapat memberikan sebuah rekomendasi ataupun informasi kepada user sehingga dapat mempermudah *user* menentukan pilihannya[7]. Sistem rekomendasi dapat memberikan sebuah rekomendasi atau informasi item kepada user melalui data yang sudah dimiliki oleh user maupun data yang sudah dimasukkan oleh pembuat, oleh karena itu diperlukan model sistem rekomendasi yang sesuai agar dapat memberikan rekomendasi atau informasi yang diperlukan oleh *user* yang menggunakannya[8].

Banyak peneliti telah melakukan penelitian terkait sistem rekomendasi menggunakan metode *collaborative filtering* di dalamnya. Penelitian [3] moh. Irfan menjelaskan sistem rekomendasi yang dibangun menggunakan Teknik *collaborative filtering* relatif akurat dengan nilai rata-rata MAE = 1,064. Pada penelitian yang dilakukan oleh hafiyah abdul Aziz[9] melakukan perancangan sistem rekomendasi penginapan. Metode yang digunakan pada penelitian ini menggunakan *collaborative filtering* dan *k-nearest neighbors* yang di gunakan untuk pengelompokan data yang digunakan. Percobaan dilakukan sebanyak 4kali dengan $k = 5$ berdasarkan pengujian tersebut mendapatkan akurasi terbaik yang didapatkan adalah 87,79% dan MAE sebesar 0,4885. Pada penelitian yang dilakukan oleh Phongsavanh Phorasim[10] melakukan perancangan sistem rekomendasi film. Metode yang digunakan pada penelitian ini menggunakan *collaborative filtering* untuk menentukan rekomendasi itemnya dan *k-means* sebagai pengelompokan item dari dataset yang dimiliki. Pengujian dilakukan dengan jumlah $K(\text{kluster}) = 10$ berdasarkan cimilarity dari rating user. Rating dibagi menjadi 5 poin di mana sangat disukai mendapat point 5, suka mendapat point 4, normal mendapat point 3, tidak suka mendapat point 2, dan sangat tidak suka mendapatkan point 1 mendapatkan kesimpulan *collaborative filtering* merupakan metode yang cukup sukses dalam memberikan rekomendasi kepada user.

Penelitian ini bertujuan untuk membangun sistem rekomendasi anime dengan menggunakan metode *collaborative filtering* baik item based ataupun item based. Metode *collaborative filtering* dipilih dikarenakan banyak penelitian menyimpulkan

bahwa metode *collaborative filtering* memiliki keakuratan yang bagus untuk sistem rekomendasi.

A. Topik dan Batasannya

Dalam penelitian ini akan dibahas bagaimana mengimplementasikan metode *collaborative filtering* dalam membangun sistem rekomendasi anime dengan harapan menghasilkan hasil rekomendasi yang baik. Adapun dataset yang di gunakan adalah dataset Anime yang didapatkan dari website www.kaggle.com[11] yang berisikan 73,516 user dan 12,294 anime. Batasan pembahasan dari pembuatan tugas akhir ini membangun sistem rekomendasi menggunakan Teknik *collaborative filtering* baik item based ataupun user based dan melakukan pengecekan sistem rekomendasi menggunakan metode MAE dan NDCG.

B. Tujuan

Tujuan dari penelitian ini adalah sebagai berikut:

1. membangun sistem rekomendasi anime menggunakan *collaborative filtering* berbasis algoritma *KNNWithMeans*
2. mengukur performansi sistem menggunakan MAE dan NDCG

C. Organisasi Tulisan

Pada bab 2 membahas mengenai studi terkait yang berkaitan dengan penelitian ini. Pada bab 3 membahas mengenai teori-teori yang digunakan dan rancangan serta alur sistem penelitian. Pada bab 4 membahas mengenai hasil dan analisis terkait penelitian ini dan pada bab 5 membahas mengenai kesimpulan berdasarkan analisis yang dilakukan.

II. KAJIAN TEORI

Pada penelitian sebelumnya sudah ada beberapa penelitian yang melakukan penelitian terkait sistem rekomendasi. Berbagai metode dan teori yang telah digunakan dalam melakukan penelitian terkait sistem rekomendasi untuk mendapatkan hasil yang akurat. Berikut merupakan beberapa penelitian yang berkaitan dengan metode yang diusulkan pada penelitian ini.

Pada penelitian yang dilakukan oleh Zhubatkhani [5] melakukan percobaan perbandingan model sistem rekomendasi pada rekomendasi film menggunakan metode *collaborative filtering* beberapa model yang digunakan adalah *NormalPredictor*, *BaselineOnly*, *KNNBasic*, *KNNWithMeans*, *KNNBaseline*, *SVD*, dan *NMF* sedangkan menggunakan metode MAE dan RMSE untuk pengukuran performa dari Model yang digunakan dari penelitian tersebut didapatkan kesimpulan metode *SVD* dan *KNNWithMeans* mendapatkan Nilai MAE terbaik yaitu sekisaran (0,7528 terbesar dan 0,7347 terkecil) sedangkan metode *SVD* dan *SlopeOne* mendapatkan Nilai RMSE terbaik yaitu sekisaran (0,9373 terkecil dan 0,9493 terbesar). Pada penelitian yang dilakukan oleh Jesper Hedlund[4] melakukan percobaan sistem

rekomendasi buku dan author menggunakan metode yang berbeda beda seperti *collaborative filtering*, *content-based filtering*, ataupun dengan metode *hybrid* dengan algoritma yang berbeda beda seperti *K-Nearest Neighbour(KNN)*, *K-Nearest Neighbour(KNN) whit means*, *K-Nearest Neighbour(KNN) with means and z-score* dan lain sebagainya metode pengukuran yang di gunakan untuk menguji sistem yang di buat adalah dengan mengecek nilai MAE dan RMSE dari setiap metode yang di gunakan dari hasil penlitian yang di lakukan mendapatkan nilai MAE = 0.7468 dan Nilai RMSE= (1.1576). Sedangkan Pada penelitian yang dilakukan oleh Hafiyah Abdul Aziz[9] melakukan perancangan sistem rekomendasi penginapan. Metode yang digunakan pada penelitian ini menggunakan *collaborative filtering* dan *k-nearest neighbors* yang di gunakan untuk pengelompokan data yang digunakan. Percobaan dilakukan sebanyak empat kali dengan $k = 5$ berdasarkan pengujian tersebut mendapatkan akurasi terbaik yang didapatkan adalah 87,79% dan MAE sebesar 0,4885.

Pada penelitian yang dilakukan oleh Phongsavanh Phorasim[10] melakukan perancangan sistem rekomendasi film. Metode yang digunakan pada penelitian ini menggunakan *collaborative filtering* untuk menentukan rekomendasi itemnya dan *k-means* sebagai pengelompokan item dari dataset yang dimiliki. Pengujian dilakukan dengan jumlah $K(\text{kluster}) = 10$ berdasarkan cimelarity dari rating user. Rating dibagi menjadi 5 rating dimana sangat disukai mendapat rating 5, suka mendapat rating 4, normal mendapat rating 3, tidak suka mendapat rating 2, dan sangat tidak suka mendapatkan point 1 mendapatkan kesimpulan *collaborative filtering* merupakan metode yang cukup sukses dalam memberikan rekomendasi kepada user.

Pada penelitian yang dilakukan oleh Tukino[12] melakukan pengoptimalan data nilai mahasiswa yang sebelumnya belum optimal agar bisa membantu mahasiswa menentukan tugas akhir. Metode yang digunakan pada penelitian ini menggunakan *k-mean* untuk menentukan grup pada data yang dimiliki selain menggunakan *k-mean* penelitian ini menggunakan aplikasi rapidminer yang bertujuan untuk data preprocessing dan visualisasi. Hasil yang dihasil dari penelitian ini adalah jumlah anggota setiap grup yang sudah dipisahkan.

Pada penelitian yang di lakukan oleh Riyan alfa rizkie[13] melakukan perancangan sistem rekomendasi wisata kuliner kota Palembang. Metode yang digunakan pada penelitian ini menggunakan *collaborative filtering* sebagai pemberian rekomendasi item dan *k-means* yang digunakan untuk membuat kelompok dari dataset yang dimiliki, selain itu di penelitian ini juga menggunakan metode recall dan precision yang berguna untuk menguji apakah rekomendasi yang diberikan oleh sistem sudah sesuai dengan kebutuhan pengguna. Dari hasil penelitian yang dilakukan mendapatkan kesimpulan

rekomendasi yang dibuat berhasil dikarenakan mendapatkan nilai dari recall dan precision sebesar 15,556% dan 70% selain itu juga mendapatkan nilai 80% relevansi dari pengguna.

Pada penelitian yang dilakukan oleh Phonexay Vilakone[14] melakukan perancangan sistem rekomendasi film berdasarkan informasi personal user. Metode yang digunakan menggunakan *k-clique* sebagai bantuan untuk mengolah rekomendasi untuk user dan Normalized Discounted Cumulative Gain digunakan untuk memberikan nilai rekomendasi dari user tersebut dari penelitian disini tingkat akurasi dari sistem sebesar 87.28%.

Pada penelitian yang dilakukan oleh Rudy prabowo[15] melakukan perancangan sistem rekomendasi film metode yang digunakan dipenelitian ini adalah pairwise preference regression sebagai sistem rekomendasi dan menggunakan *Normalized Discounted Cumulative Gain* untuk melihat hasil akurasi sistem yang dibuat dari penelitian ini berhasil dikarenakan mendapat nilai terbaik Normalized Discounted Cumulative Gain (nDCG) 0.8484.

A. Preprocessing Data

Preprocessing data merupakan sebuah metode untuk membuat data menjadi lebih mudah di proses. Tujuan dari preprocessing data untuk meningkatkan kualitas data termasuk akurasi, konsistensi kelengkapan dan interpretabilitas dari sebuah data[16].

B. Collaborative filtering

Collaborative filtering merupakan salah satu metode dalam rekomendasi sistem. *collaborative filtering* memberikan rekomendasi item berdasarkan riwayat pengguna dalam menyukai, menilai, dan menggunakan sesuatu item yang pernah digunakan[9]. Dasar asumsi dari metode *collaborative filtering* bahwa user dapat dikelompokkan dalam selera yang sama mereka membuat asumsi bahwa jika orang yang memiliki selera yang sama maka orang dalam kelompok tersebut juga menyukai item tersebut meskipun belum membeli atau menggunakannya[10].

Kualitas dari rekomendasi yang dihasilkan oleh metode *collaborative filtering* sangat tergantung dari penilaian item baru user yang lain. *Collaborative filtering* pada dasarnya dibagi menjadi dua kategori yaitu:

1. Item base metode menerapkan perhitungan kesamaan item- item yang sudah dirating dengan item- item lain Langkah berikutnya dipilih item- item yang memiliki nilai kemiripan dengan item yang sudah dirating. Nilai kemiripan tersebut disimpan lalu dijadikan nilai ukur buat memprediksi nilai rating pada item yang ditunjukkan[17].

TABEL 1
DATA RATING ANIME

	naruto	clanat	One piace
adi	2		3
bayu	5	2	
cici	3	3	1
danny		2	2

Pada table di atas *user* di atas yang akan di prediksi adalah 3 anime yang masih kosong dari setiap *user*, untuk mendapatkan rating yang kosong pertama tama mencari *similarity* atara ke tiga anime tersebut.

TABEL 2
DATA SIMILARITY ANTAR ANIME

Sim(Naruto,clanat)	0.904
Sim(clanat,one piace)	0.868
Sim(Naruto, one piace)	0.789

Dari data tabel di atas telah mendapatkan *similarity* atar ke tiga anime tersebut. Langkah berikutnya adalah dengan melakukan prediksi untuk rating rating kosong dari ke tiga anime.

TABEL 3
DATA PREDIKSI RATING ANIME

Prediksi(adi,clanat)	2.490
Prediksi (bayu,one piace)	3.429
Prekdisi(danny, naruto)	2.00

2. *User base* metode ini mengukur kesamaan selera dengan membuat peringkat dari item yang sudah pernah dirating oleh user lain[18].

TABEL 4
DATA RATING ANIME

	naruto	clanat	One piace	overlord	To love ru
anggi	5	4	1	4	
budi	3	1	2	3	3
caca	4	3	4	3	5
doni	3	3	1	5	4

Pada tabel di atas *user* yang akan di prediksi adalah anggi dimana kita akan memprediksi rating pada anime *to love ru*. tahapan awal yang di lakukan adalah mecari pengguna yang memiliki prefrensi sama dengan anggi. Lakukan perhitungan rata-rata rating setiap pengguna untuk anime selain anime *to love ru* seperti berikut:

TABEL 5
DATA RATA-RATA RATING ANIME

	naruto	clanat	One piace	overlord	To love ru	Rata-rata rating
anggi	5	4	1	4		3.50
budi	3	1	2	3	3	2.25
caca	4	3	4	3	5	3.50
doni	3	3	1	5	4	3.00

Dari tabel di atas setelah mendapat rata-rata dari setiap *user* setelah mendapatkan rata-rata rating dari setiap user kemudian mencari selisih dari rata-rata user dengan anime yang mereka rating.

TABEL 6
DATA SELISIH RATING ANIME DENGAN RATING ASLI

	naruto	clanat	One piace	overlord	To love ru
anggi	1.50	0.50	-2.50	0.50	
budi	0.75	-1.25	-0.25	0.75	
caca	0.50	-0.50	0.50	-0.50	
doni	0.00	0.00	-2.00	-2.00	

Dari tabel di atas setelah mendapatkan selisih dari setiap anime yang dirating tahapan berikutnya adalah mencari nilai *similarity* antara anggi dengan user lainnya.

TABEL 7
DATA SIMILARITY ANTAR USER

Sim(anggi,budi)	0.302
Sim(anggi,caca)	-0.333
Sim(anggi,doni)	0.707

Dari data dabel di atas dapat di lihat nilai *similarity* setelah mendapatkan nilai *similarity* di cari nilai *similarity* yang paling besar dari tabel di atas kita dapat menentukan prediksi dengan caara menentukan Batasan kemiripan nilai *similarity* pengguna, missal batas *similarity* di bawah 0.2 tidak di hitung.

Bedasarkan hasil *similarity* berikutnya dilakukan perhitungan untuk prediksi rating anggi pada anime *to love ru* dengan menggunakan *neighbour* budi dan doni, yaitu nilai rata-rata anggi ditambahkan dengan seluruh nilai *similarity* antara pengguna berikutnya dikali selisih rating dan rerata dibagi dengan nilai absolute dari *similarity* beberapa pengguna tersebut.sehingga mendapatkan prediksi *rating* anime *to love ru* pada anggi sebesar 4.425.

C. NDCG

Normalized discounted comulative gain method (NDCG) merupakan teknik untuk mengukur kinerja dari sistem rekomendasi berdasarkan tingkat

relevansi dari rekomendasi suatu system. Hasil dari NDCG biasanya di antara nilai 0.0 sampai 1.0 dimana nilai 1.0 merupakan nilai ideal dari suatu rekomendasi[15].

Ada dua keunggulan yang didapat saat menggunakan metode ini :

1. Memungkinkan semua item yang diambil memiliki nilai yang relevance
2. Nilai relevan masih bisa didapatkan dengan data yang sedikit[14]

Rumusan yang digunakan dalam metode ini adalah :

$$NDCG = \frac{DCG}{IDCG} \quad (1)$$

$$DCG = \sum_{i=1}^p \frac{rel_i}{\log_2(i+1)} \quad (2)$$

$$IDCG = \sum_{i=1}^{|rel|} \frac{2^{rel_i} - 1}{\log_2(i+1)} \quad (3)$$

DCG = nilai dari prediksi item

IDCG = Nilai estimasi item yang di prediksi

P = akumulasi posisi peringkat suatu data

rel= daftar data yang relevan dalam korpus data

D. Mean Absolute Error(MAE)

Mean Absolute Error(MAE) merupakan salah satu metode yang digunakan untuk mengukur error dari prediksi rating yang akan dihasilkan oleh sistem rekomendasi, Semakin kecil nilai MAE yang dihasilkan maka semakin baik juga semakin bari sistem rekomendasi itu dibuat[19].

Mean Absolute Error(MAE) bisa disebut juga sebuah matric yang biasanya digunakan untuk membandingkan algoritma pemrosesan sinyal. Oleh karena itu memerlukan sinyal referensi. Namun jika sinyal referensi tidak ada atau tidak dapat diukur, maka MAE tidak terdefinisi berikut rumus persamaannya[20]:

$$MAE = \frac{1}{N} \sum_{i=0} |R_i - P_i| \quad (4)$$

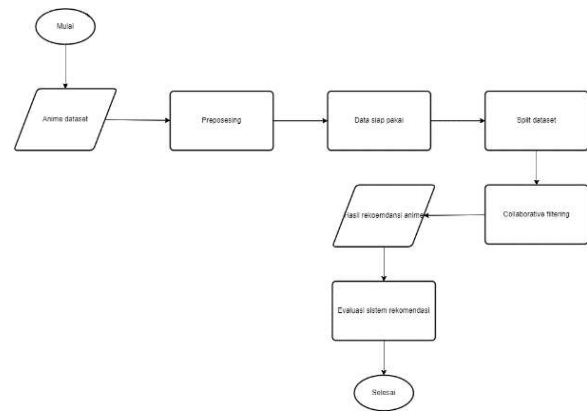
N = jumlah semua item yang memiliki rating prediksi

R_i = rating sebenarnya dari item yang di gunakan

P_i = nilai estimasi dari item yang di prediksi

III. METODE

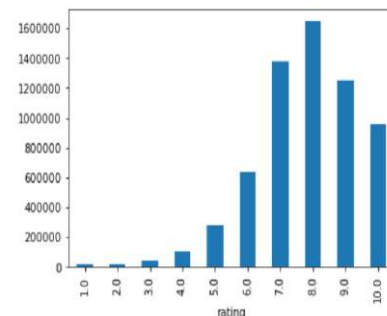
Pada penelitian ini membangun sebuah sistem rekomendasi anime dengan menggunakan metode *Collaborative filtering*. Adapun alur sistem yang akan dibangun dapat dilihat pada gambar berikut.



GAMBAR 1
ALUR SISTEM YANG DIBUAT

A. Dataset

Data ini yang di dapatkan dari www.kaggle.com[11]. Terdapat dua data yang didapatkan dari website tersebut yang pertama data anime yang berisikan 12.294 data yang memiliki kolom yaitu anime id, nama yang berisikan judul anime, genre, type, episode, rating, dan members. Data rating yang memiliki beberapa kolom yaitu berisikan rating yang merupakan hasil penilaian user dari anime yang pernah di totol oleh user, anime id, dan user id yang diberikan user sebesar 7.813.737 data dengan statistic data sebagai berikut



GAMBAR 2
DATA RATING USER

Dari gambar data di atas dapat di simpulkan rata rata user banyak yang memberi rating anime 8 dan yang kedua terbesar memberikan rating 7.

B. Preprocessing Data

Dalam Penelitian ini melakukan beberapa Preprosesing agar data yang diggunakan dapat digunakan dalam penelitian ini. Yang pertama adalah data *Cleaning*, Data cleaning merupakan tahapan dimana membersihkan data yang tidak akan di gunakan pada penelitian yang dilakukan. Pada tahapan ini data rating yang diggunakan memiliki rating yang tidak memiliki nilai atau (NaN) dapat di lihat pada tabel 8 di bawah hal ini diperlukan penghilangan data yang rating yang tidak memiliki nilai(NaN) agar data dapat di gunakan hasil dari penghilangan NaN dapat di lihat pada table 9 dimana user 1 yang memiliki rating NaN di hapus dari tabel.

TABEL 8
DATA RATING SEBELUM DATA CLEANING

user_id	anime_id	rating
---------	----------	--------

1	20	NaN
3	20	8.0
5	20	6.0

TABEL 9
DATA HASIL DATA CLEANING

user_id	anime_id	rating
3	20	8.0
5	20	6.0
21	20	8.0

yang kedua melakukan data *Integration* yaitu dimana menggabungkan kedua data yang digunakan agar menjadi satu data seperti pada table 10 data awalnya berisikan anime_id, nama anime, genre, type, episode, rating, dan Member dengan melakukan

TABEL 11
DATA HASIL INTEGRATION

User_id	anime_id	Rating_user	name	genre	type	episodes	rating	Member
1	8074	10.0	Highschool of the Dead	Action	TV	12	7.46	535892
3	8074	6.0	Highschool of the Dead	Action	TV	12	7.46	535892
5	8074	2.0	Highschool of the Dead	Action	TV	12	7.46	535892

yang ke ketiga data *Transformation* yaitu dimana mengubah data yang digunakan agar sesuai dengan kebutuhan dalam penelitian yang dilakukan. pada data tabel 12 terdapat kolom user id, anime id, rating user, nama anime, genre type, episode, rating, dan member Kolom yang di perlukan agar bisa di

data integration dataset sebelumnya di gabungkan dengan data rating user berdasarkan anime hal ini dapat di lihat pada table 11 di mana data awal memiliki tambahan kolom berupa user id dan rating user.

TABEL 10
DATA USER SEBELUM INTEGRATION

anime_id	name	genre	type	episode	rating	Member
32281	Kimino Na wa.	Drama	Moviem	1	9.37	200630
5114	Full metal Alchemist : Brotherhood	Action	TV	64	9.26	793665
28977	Gintama	Action	TV	51	9.25	114262

gunakan dalam penelitian ini adalah user id, anime id rating user dan nama yang berisikan nama anime yang telah di rating oleh user lalu rating user di rubah nama kolomnya menjadi rating, pada tabel 13 dapat dilihat hasil akhir dari data yang di gunakan adalah data yang berisikan user id, anime id, rating, dan nama anime.

TABEL 12
DATA SEBELUM TRANSFORMATION

User_id	anime_id	Rating_user	name	genre	type	episodes	rating	Member
1	8074	10.0	High school of the Dead	Action	TV	12	7.46	535892
3	8074	6.0	High school of the Dead	Action	TV	12	7.46	535892
5	8074	2.0	High school of the Dead	Action	TV	12	7.46	535892

TABEL 13
DATA SESUDAH TRANSFORMATION

user_id	anime_id	rating	name
3	20	8.0	Naruto

5	20	6.0	Naruto
21	20	8.0	Naruto

yang ke empat melakukan data *Reduction* yaitu mengurangi data yang digunakan agar bisa

digunakan kedalam penelitian yang di lakukan di mana data awal sebanyak 6.337.241 data setelah melakukan data *Reduction* dilakukan menyadi sebanyak 1.674.551 data.

C. Cross Validation

Cross validation merupakan salah satu metode pembagian sebuah dataset menjadi data latih dan data uji.

TABEL 14
CROSS VALIDATION DENGAN K-FOLD = 5

Cross Validation	Dataset yang dibagi menjadi 5 bagian	Percobaan
Dataset	Data 1	Percobaan-1
	Data 2	Percobaan-2
	Data 3	Percobaan-3
	Data 4	Percobaan-4
	Data 5	Percobaan-5

Untuk membagi dataset ke dalam data *train* dan data *test* cross validation akan menggunakan *k-fold* yang terdiri atas beberapa bagian. apabila *k-fold* yang digunakan sebanyak 5, maka dataset akan dibagi menjadi 5. Sehingga dataset akan terdiri atas empat data *train* dan satu data *test*. Tabel 14 merupakan gambaran dataset yang dipecah menggunakan cross validation dengan *k-fold* = 5 dimana kotak berwarna hitam menandakan data *test* dan kotak berwarna abu-abu menunjukkan data *train*.

D. Collaborative Filtering

Setelah dataset dipisah data yang di gunakan untuk melakukan collaborative filtering adalah kolom yang berisikan user id, anime id ,dan rating untuk menghitung dengan persamaan user based dan item based yang di mana user based metode ini mengukur kesamaan selera dengan membuat peringkat dari item yang sudah pernah dirating oleh user lain dan item based melakukan perhitungan kemiripan item-item yang sudah dirating dengan item-item lainnya Langkah berikutnya dipilih item-item yang mempunyai nilai yang mendekati dengan item yang telah dirating. Nilai yang mendekati dengan item yang telah di rating tersebut disimpan lalu dijadikan nilai ukur untuk memprediksi nilai rating pada item yang ditujukan dengan rumus sebagai berikut:

$$\hat{r}_{ui} = \mu_u + \frac{\sum_{v \in N_i^k(u)} \text{sim}(u, v) \cdot (r_{vi} - \mu_v)}{\sum_{v \in N_i^k(u)} \text{sim}(u, v)} \quad (5)$$

Atau

$$\hat{r}_{ui} = \mu_i + \frac{\sum_{u \in N_i^k(i)} \text{sim}(i, j) \cdot (r_{uj} - \mu_j)}{\sum_{u \in N_i^k(i)} \text{sim}(i, j)} \quad (6)$$

Di mana $\hat{r}_{ui}$ merupakan prediksi μ_u merupakan mean *user* u dan similarly μ_i merupakan mean dari item i, k merupakan jumlah ketetanggan, $\text{sim}(u, v)$ dan $\text{sim}(i, j)$ untuk memastikan literasi keberapa *collaborative filtering* di lakukan. r_{vi} merupakan rating dari user v dalam item i dan similarly r_{uj} merupakan rating user u dalam item j, μ_v merupakan mean dari user v dan similarly μ_j merupakan mean dari item j, [4], [5], [21].

E. Melakukan Evaluasi Hasil Rekomendasi Akhir

Pengujian sistem rekomendasi yang dibangun menggunakan *collaborative filtering* dilakukan pengecekan dengan metode *MAE* dan *NDCG* untuk *MAE* nilai yang di gunakan rating asli dari data yang digunakan dan dibandingkan dengan nilai estimasi yang didapat dari *collaborative filtering* sedangkan untuk *ndcg* menggunakan rating lalu di bandingkan dengan nilai estimasi yang didapat dari *collaborative filtering*.

IV. HASIL DAN PEMBAHASAN

Pada penelitian ini, pertama data rating yang memiliki nilainya negatif di ubah menjadi NaN yang kedua data rating dan data user digabungkan menjadi satu berdasarkan *user id* lalu data user yang digunakan adalah sebanyak 20.000 *user* , yang ke tiga data yang sudah digabungkan diambil kolom yang berisikan data *user id*, *anime id*, nama *anime* lalu menghapus data yang telah di rubah menjadi NaN. data hasil *preprocessing* berikutnya akan digunakan untuk membangun sistem rekomendasi menggunakan metode *colaborativ filtering berbasis KNNwithMeans*. sistem yang dibangun akan dievaluasi dengan menggunakan beberapa skenario pengujian untuk mengetahui kombinasi terbaik sehingga menghasilkan performansi yang paling bagus. Berikut ini beberapa skenario pengujian yang digunakan untuk evaluasi model:

A. skenario pengujian perbandingan dengan menggunakan cross validation

Pada scenario ini data yang sudah melewati *preprocessing* dimasukan kedalam *cross validadion* dengan jumlah percobaan lima kali percobaan dan sepuluh kali percobaan dengan menggunakan metode *collaborativ filtering* dengan model algoritma *KNNwithMean*.

TABEL 15
PENGUJIAN PERBANDINGAN DENGAN MENGGUNAKAN CROSS VALIDATION

Cross validadion	Jumlah k(ketetanggaan)	MAE terbaik k	Rata-rata MAE	NDCG terbaik	Rata rata NDCG
5	5	1.0241	1.0266	0.2022	1.0045
10	5	1.0199	1.0233	0.1785	1.7548

Bedasarkan hasil pengujian di atas didapatkan Nilai MAE terbaik = 1.0199 pada table 15 dengan rasio cross validation = 10 sedangkan untuk nilai NDCG terbaik terdapat pada table 15 dengan rasio cross validation = 5 dengan nilai NDCG = 0.2022. Oleh sebab itu dapat disimpulkan bahwa jumlah cross validation dapat mempengaruhi besar kecilnya MAE dan NDCG yang di dapatkan.

B. Skenario pengujian perbandingan jumlah

K(ketetanggan) pada algoritma *KNNwithMeans*

Pada scenario ini data yang sudah melewati preprocessing dimasukan kedalam *cross validadion* dengan jumlah percobaan lima kali percobaan dan sepuluh kali percobaan dengan menggunakan metode collaborativ filtering dengan model algoritma *KNNwithMeans* dengan jumlah k(ketetanggan) = 5 dan 50.

TABEL 16
PERBANDINGAN JUMLAH K(KETETANGGAN) PADA
ALGORITMA *KNNWHITMEANS*

Cross valita dion	Jumlah k(keteta nggaan)	MAE terbaik	Rata- rata MAE	NDCG terbaik	Rata rata NDCG
5	5	1.0241	1.0266	0.2022	1.0045

TABEL 17
PENGUJIANDENGAN MENGGUKAAN ALGORITMA *KNNWHITMEANS* DENGAN METODE USER BASED DAN ITEM BASED

Cross valita dion	Jumlah k(ketetan ggaan)	Item based/ user based	MAE terbaik	Rata- rata MAE	NDCG terbaik	Rata rata NDCG
5	5	Item based	1.0241	1.0266	0.2022	1.0045
5	5	user based	0.9626	0.9644	0.2012	1.0032
5	50	Item based	0.9421	0.9433	0.2027	1.0065
5	50	user based	0.9056	0.9062	0.2028	1.0062
10	5	Item based	1.0199	1.0233	0.1785	1.7548
10	5	user based	0.9512	0.9546	0.1767	1.7506
10	50	Item based	0.9368	0.9400	0.1772	1.7574
10	50	user based	0.8989	0.9016	0.1769	1.7599

Bedasarkan hasil pengujian di atas didapatkan Nilai MAE terbaik = 0.8989 pada table 17 dengan rasio cross validation = 10 dan K(ketenggan) *KNNwithMean* = 50 (user based) sedangkan untuk nilai NDCG terbaik terdapat pada rasio cross validation = 5 dan K(ketenggan) *KNNwithMean* = 50 (user based) dengan nilai NDCG = 0.2027. Oleh sebab itu dapat disimpulkan bahwa jumlah cross validation, K(ketenggan) *KNNwhithMean*, dan dengan metode *item based* atau *user based* dapat mempengaruhi besar kecilnya MAE dan NDCG yang di dapatkan.

5	50	0.9421	0.9433	0.2027	1.0065
10	5	1.0199	1.0233	0.1785	1.7548
10	50	0.9368	0.9400	0.1772	1.7574

Bedasarkan hasil pengujian di atas didapatkan Nilai MAE terbaik = 0.9368 pada table 16 dengan rasio cross validation = 10 dan K(ketenggan) *KNNwithMean* = 50 sedangkan untuk nilai NDCG terbaik dengan rasio cross validation = 5 dan K(ketenggan) *KNNwithMean* = 50 dengan nilai NDCG = 0.2027. Oleh sebab itu dapat disimpulkan bahwa jumlah cross validation dan K(ketenggan) *KNNwhithMean* dapat mempengaruhi besar kecilnya MAE dan NDCG yang di dapatkan.

C. Skenario pengujian dengan menggunakan algoritma *KnnwhitMeans* dengan metode user based dan item based

Pada scenario ini data yang sudah melewati preprocessing dimasukan kedalam *cross validadion* dengan jumlah percobaan lima kali percobaan dan sepuluh kali percobaan dengan menggunakan metode collaborativ filtering dengan model algoritma *KNNwithMean* dengan jumlah k(cluster) = 5 dan 50 serta menggunakan metode item based dan user based.

V. KESIMPULAN

Penulisan tugas akhir ini bertujuan pada membangun sistem rekomendasi anime menggunakan metode collaborative filtering serta melihat seberapa akurat sistem rekomendasi yang di bangun dengan melakukan pengecekan nilai MAE dan NDCG berdasarkan pengujian yang telah dilakukan terhadap metode yang di gunakan didapatkan kesimpulan sebagai berikut:

- Hasil pengujian dari sistem rekomendasi yang dibangun menggunakan metode *collaborative filtering* mendapatkan nilai MAE terbaik sebesar 0.8989 dan nilai NDCG terbaik sebesar 0.2027

- B. Jumlah pembagian *cross validadion* berpengaruh dengan hasil MAE dan NDCG pada sistem rekomendasi yang di bangun
- C. Jumlah k(ketetanggan) berpengaruh pada hasil MAE dan NDCG pada sistem rekomendasi yang dibangun
- D. Metode *User based* dan *Item based* mempengaruhi nilai MAE dan NDCG dikarenakan tergantung dari user yang direkomendasikan memiliki kecocokan yang sama dengan user yang lain atau tidak

Saran yang dapat dipertimbangkan untuk penelitian selanjutnya adalah mencoba metode rekomendasi lain atau pun bisa mingguankan metode *hybird* untuk melakukan perbandingan dari sistem rekomendasi yang sudah dibuat.

REFERENSI

- [1] Nuurshadieq and A. T. Wibowo, "Leveraging Side Information to Anime Recommender System using Deep learning," in 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, Dec. 2020, pp. 62–67. doi: 10.1109/ISRITI51436.2020.9315363.
- [2] "Top Anime (18350 -)," MyAnimeList.net. <https://myanimelist.net/topanime.php?limit=18350> (accessed Aug. 20, 2021).
- [3] M. Irfan, "SISTEM REKOMENDASI: BUKU ONLINE DENGAN METODE COLLABORATIVE FILTERING," vol. 7, no. 1, p. 9, 2014.
- [4] J. Hedlund and E. N. Tengstrand, "A Comparison between Different Recommender System Approaches for a Book and an Author Recommender System," p. 97.
- [5] A. Y. Zhubatkhani, Z. A. Buribayev, S. S. Aubakirov, M. D. Dilmagambetova, and S. A. Ryskulbek, "COMPARISON MODELS OF MACHINE LEARNING FOR MOVIE RECOMMENDATION SYSTEMS," News of NAS RK. SPh-M, vol. 335, no. 1, pp. 26–31, Feb. 2021, doi: 10.32014/2021.2518-1726.4.
- [6] A. Lesley, "What Is Anime? - Anime." <http://www.bellaonline.com/articles/art4260.asp> (accessed Jul. 27, 2022).
- [7] A. Halim, H. Gohzali, D. M. Panjaitan, and I. Maulana, "Sistem Rekomendasi Film menggunakan Bisecting K-Means dan Collaborative Filtering," p. 5, 2017.
- [8] S. Sari and A. P. Sary, "SISTEM REKOMENDASI PERSONAL PADA TOKO BUKU ONLINE MENGGUNAKAN PENDEKATAN COLLABORATIVE FILTERING DAN ALGORITMA SLOPE ONE," p. 6, 2017.
- [9] H. A. Aziz, B. Dirgantoro, and C. Setianingsih, "Boarding Recommendation Application Using Collaborative Filtering Algorithm," p. 11, 2021.
- [10] P. Phorasim and L. Yu, "Movies recommendation system using collaborative filtering and k-means," IJACR, vol. 7, no. 29, pp. 52–59, Feb. 2017, doi: 10.19101/IJACR.2017.729004.
- [11] "Anime Recommendations Database." <https://kaggle.com/CooperUnion/anime-recommendations-database> (accessed Oct. 04, 2021).
- [12] T. Tukino and B. Huda, "PENERAPAN ALGORITMA K-MEANS UNTUK Mendukung Keputusan Dalam Pemilihan Tema Tugas Akhir Pada Prodi Sistem Informasi Universitas Buana Perjuangan Karawang.," TX, vol. 4, no. 1, pp. 1–10, Apr. 2019, doi: 10.36805/technoexplo.v4i1.542.
- [13] R. A. Rizkie and M. Fachrurrozi, "Sistem Rekomendasi Wisata Kuliner Kota Palembang Menggunakan Metode Collaborative Filtering," p. 3.
- [14] K. X. Phonexay Vilakone and Doo-Soon Park, "Movie Recommendation System Based on Users' Personal Information and Movies Rated Using the Method of k-Clique and Normalized Discounted Cumulative Gain," Journal of Information Processing Systems, vol. 16, no. 2, pp. 494–507, Apr. 2020, doi: 10.3745/JIPS.04.0169.
- [15] R. Rismala, R. Prabowo, and A. T. Wibowo, "Pairwise Preference Regression on Movie Recommendation System," IndoJC, vol. 4, no. 1, p. 57, Mar. 2019, doi: 10.21108/INDOJC.2019.4.1.255.
- [16] S. Agarwal, "Data Mining: Data Mining Concepts and Techniques," in 2013 International Conference on Machine Intelligence and Research Advancement, Katra, India, Dec. 2013, pp. 203–207. doi: 10.1109/ICMIRA.2013.45.
- [17] V. L. Jaja, B. Susanto, and L. R. Sasongko, "Penerapan Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Data MovieLens," dC, vol. 9, no. 2, p. 78, Sep. 2020, doi: 10.35799/dc.9.2.2020.28274.
- [18] V. X. Chen and T. Y. Tang, "Incorporating Singular Value Decomposition in User-based Collaborative Filtering Technique for a Movie Recommendation System: A Comparative Study," in Proceedings of the 2019 the International Conference on Pattern Recognition and Artificial Intelligence - PRAI '19, Wenzhou, China, 2019, pp. 12–15. doi: 10.1145/3357777.3357782.
- [19] E. Vozalis and K. G. Margaritis, "Analysis of Recommender Systems' Algorithms," p. 15.

- [20] T. McMurray and J. A. Pearce, "Theoretical and experimental comparison of the Lorenz information measure, entropy, and the mean absolute error," in Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation, Dallas, TX, USA, 1994, pp. 24–29. doi: 10.1109/IAI.1994.336688.
- [21] "k-NN inspired algorithms — Surprise 1 documentation."
https://surprise.readthedocs.io/en/stable/knn_inspired.html (accessed Sep. 09, 2022).

