

## Pembuatan Sistem Rekomendasi Film dengan Menggunakan Metode Item Based Collaborative Filtering pada Apache Mahout

Kiki Ratna Sari<sup>1</sup>, Wildan Suharso<sup>2</sup>, Yufiz Azhar<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika/Universitas Muhammadiyah Malang

kikiratnasari9@gmail.com<sup>1</sup>, wsuharso@umm.ac.id<sup>2</sup>, yufiz@umm.ac.id<sup>3</sup>

### Abstrak

Sistem rekomendasi merupakan sebuah metode yang digunakan untuk memberikan rekomendasi pada sebuah produk seperti buku, musik dan film dengan memberikan nilai prediksi tertinggi pada penggunaannya. Sistem rekomendasi saat ini masih bisa ditingkatkan lagi kualitasnya, sehingga dengan menggunakan metode baru diharapkan dapat lebih meningkatkan lagi nilai relevansi dari hasil rekomendasi yang diberikan daripada sistem-sistem sebelumnya. Selain itu perkembangan film didunia juga setiap harinya jumlahnya semakin meningkat dengan berbagai jenis genre yang dimiliki, sehingga membuat para penonton film merasa kesulitan untuk memilih film apa yang akan ditonton. Maka dari itu, diperlukan adanya sebuah sistem yang dapat memberikan rekomendasi dari penonton film lainnya. Pada penelitian ini pengembangan sistem rekomendasi film akan menggunakan metode Item Based Collaborative Filtering yang nantinya akan dikembangkan pada framework Apache Mahout dan proses pengujiannya menggunakan algoritma k-fold cross validation. Hasil dari penelitian ini menunjukkan bahwa metode Item based Collaborative Filtering memberikan hasil rekomendasi yang sangat mendekati dengan preferensi nilai yang diberikan oleh penggunaannya, hal ini ditunjukkan pada hasil pengujian sistem mendapatkan nilai akurasi sebesar 97%.

**Kata Kunci:** Sistem Rekomendasi, Item Based, Collaborative Filtering, Apache Mahout, Film

### Abstract

A recommendation system is a method used to provide recommendations on a product such as books, music and movies by providing the highest user predictive value. The current recommendation system can still be improved by the quality, so using new methods is expected to further increase the relevance value of the results of the recommendations given from previous systems. In addition to the development of film in the world as well as the increasing number of genres that have different genre, so that the movie audience feel the difficulty to choose what movie to watch. Therefore, there is a system that can provide recommendations from other film audiences. In this research the development of the film recommendation system will use the Item Based Collaborative Filtering method which will later be developed in the Apache Mahout framework and the testing process using the K-fold cross validation algorithm. The results of this research show that the Item-based Collaborative Filtering method provides a recommendation result that is very close to the preference value given by its users, it is shown in the test results of the system getting an accuracy value of 97%.

**Keywords:** Recommendation System, Item Based, Collaborative Filtering, Apache Mahout, Movie

### 1. Pendahuluan

Perkembangan teknologi informasi dan telekomunikasi didunia mengalami peningkatan yang sangat tajam, hal ini dapat dilihat bahwa sekarang ini banyak sekali kegiatan manusia yang membutuhkan teknologi informasi maupun komunikasi, tidak terkecuali dalam bidang musik maupun film. Film merupakan audio visual yang memiliki banyak sekali genre, dari genre komedi, drama, horor, *action* dan masih banyak lagi lainnya. Dikarenakan banyaknya jumlah film yang tayang didunia, dan setiap orang memiliki genre favorit sendiri-sendiri, maka dari itu dibutuhkan sebuah rekomendasi dari penonton film lainnya dengan melihat rating film apa yang menjadi pilihan terbanyak.

Sistem rekomendasi merupakan sebuah metode yang digunakan untuk memberikan rekomendasi dengan memprediksi nilai sebuah *item* bagi seorang pengguna dan kemudian

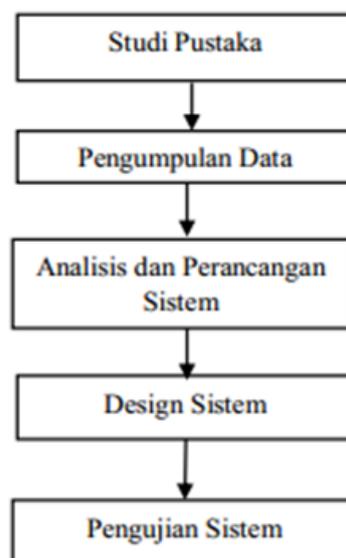
mempresentasikan *item* dengan nilai prediksi tertinggi. Prediksi dalam sistem rekomendasi dapat digunakan pada berbagai bidang yaitu meliputi buku, musik, film, tempat wisata dan masih banyak lainnya [1]. Terdapat dua pendekatan utama dalam proses pembuatan sistem rekomendasi yaitu *content based filtering* dan *collaborative filtering*. *Content based filtering* memberikan rekomendasi *item* yang kemungkinan disukai oleh *user* berdasarkan karakteristik dari *item* yang sebelumnya mendapatkan penilaian positif dari *user* yang bersangkutan. *Collaborative filtering* merupakan proses penyaringan atau pengevaluasian item dengan menggunakan opini dari orang lain [2].

Seperti penelitian yang dilakukan oleh Sumarlin [3] aplikasi sistem rekomendasi film sesuai dengan preferensi genre yang ditentukan oleh *user* berhasil dibuat dengan menggunakan metode *Simple Additive Weighting* pada *platform* Android dengan nilai *System Usefulness* sebesar 85,4% sehingga aplikasi ini sangat membantu pengguna dalam memilih film. Sari [4] dengan penelitiannya berikut juga menjelaskan bahwa dengan menggunakan *Collaborative Filtering* bersama dengan *Euclidean Distance* sebagai pengukur kemiripan rating *user* dapat memberikan rekomendasi bagi pelanggan toko film online dengan hasil evaluasi menggunakan *Kendall Rank Correlation Coefficient* menghasilkan nilai 0,66 atau 66%, hal ini menunjukkan hasil rekomendasi yang diberikan sesuai dengan keinginan pelanggan. Begitu juga dengan penelitian yang dilakukan oleh Marthasari [5] Sistem rekomendasi dibangun menggunakan aturan-aturan yang dihasilkan oleh Algoritma Apriori. Untuk dapat menampilkan barang rekomendasi digunakan nilai support 20, sedangkan nilai confidence digunakan untuk menentukan N-teratas barang untuk direkomendasikan.

Sistem rekomendasi saat ini masih dapat ditingkatkan lagi kualitasnya, yaitu dengan meningkatkan lagi nilai relevansi hasil rekomendasi yang diberikan dari penelitian sebelumnya. Berdasarkan uraian yang disampaikan diatas, maka pada penelitian kali ini akan dibangun sebuah sistem rekomendasi film dengan menggunakan metode *Item Based Collaborative Filtering* dengan *Apache Mahout* sebagai frameworknya dan proses pengujian nantinya menggunakan algoritma *k-fold cross validation*. Dengan diterapkannya metode tersebut pada penelitian ini dikarenakan metode tersebut dirasa sangat cocok dengan permasalahan yang ada, sehingga penulis ingin membuktikan seberapa cocok tingkat relevansi hasil rekomendasi yang dihasilkan oleh metode *Item Based Collaborative Filtering*.

## 2. Metode Penelitian

Secara garis besar hasil dari penelitian ini adalah sebuah sistem rekomendasi film yang dapat memberikan hasil rekomendasi yang sesuai dengan preferensi penggunaannya. Metode penelitian yang digunakan secara umum meliputi studi pustaka, pengumpulan data, analisis dan perancangan sistem, design sistem dan pengujian sistem yang ditunjukkan pada Gambar 1.



Gambar 1. Diagram metode penelitian

## 2.1 Studi Pustaka

Studi pustaka digunakan untuk mengetahui konsep dan menambah pemahaman dari sistem rekomendasi, metode yang digunakan, serta proses pengujian yang dilakukan pada penelitian ini. Studi pustaka diambil dari beberapa sumber, seperti jurnal, buku, konferensi dan lainnya. Terdapat beberapa penelitian yang digunakan sebagai acuan antara lain penelitian tentang sistem rekomendasi film [6], tentang metode collaborative filtering dengan beberapa aplikasinya [7], penelitian tentang metode item based collaborative filtering [8], penelitian yang menggunakan *Apache Mahout* [9], serta penelitian tentang bagaimana menerapkan sistem rekomendasi pada *Apache Mahout* [10].

## 2.2 Pengumpulan Data

Tahapan ini dilakukan untuk mengumpulkan data pada studi kasus baik berupa dokumen, dataset, ataupun data hasil observasi dan wawancara. Dataset yang dipergunakan pada penelitian ini diambil pada website MovieLens Dataset. Pada website tersebut sudah tersedia beberapa dataset terbaru dan terupdate dengan berbagai ukuran mengenai film diseluruh dunia dalam berbagai genre. Pada penelitian ini dataset yang digunakan berukuran 100k, dengan detail sebagai berikut : user.csv, movie.csv, link.csv, rating.csv. Contoh isi dari dataset akan ditunjukkan pada Tabel 1, Tabel 2, dan Tabel 3 seperti dibawah ini.

*Tabel 1. Dataset user.csv*

UserID	Nama	Umur	Gender	email	password
8	Ginger Lile	38	Female	gliile7@cloudflare.com	pMtSW6Fzk
9	Leroi Craiden	50	Male	lcraiden8@issuu.com	exvOUXdD
10	Debera Croan	48	Female	dcroan9@reddit.com	SaiZC6r
11	Siffre Andresen	16	Male	sandresena@ft.com	Jwy1qR
12	Sorcha Garza	23	Female	sgarza@printfriendly.com	GprDa2

Berdasarkan Tabel 1 diatas dataset user.csv merupakan dataset yang berisi identitas dari user-user yang telah mengisi rating film, data ini berisi 610 baris.

*Tabel 2. Dataset movie.csv*

movieID	title	genres
96	In the Midwinter (1995)	Comedy Drama
97	Hate (Haine, La) (1995)	Crime Drama
99	Heidi Fleiss: Hollywood Madam (1995)	Documentary
100	City Hall (1996)	Drama Thriller
101	Bottle Rocket (1996)	Adventure Comedy Romance

Berdasarkan Tabel 2 diatas dataset movie.csv merupakan dataset yang berisi data film yang ada, data berisi 9742 baris data. Pada penelitian ini penulis hanya menggunakan sebagian kecil dari data yang diambil pada MovieLens dikarenakan guna mempercepat proses perhitungan rekomendasi pada sistem.

*Tabel 3. Dataset link.csv*

movieID	imdbid	tmdbid
287	113097	10525
238	113028	27985
239	113198	15789
240	113303	27303
241	113089	21183

Berdasarkan Tabel 3 diatas dataset links.csv merupakan dataset yang berisi alamat link dari film yang bersangkutan pada IMDB.com dan TMDb.org, yang digunakan untuk mengarahkan pengguna ke website dari film yang telah diberikan rekomendasi. Data ini berisi 9742 baris data sesuai dengan data yang ada pada dataset movie.csv.

Tabel 4. Dataset rating.csv

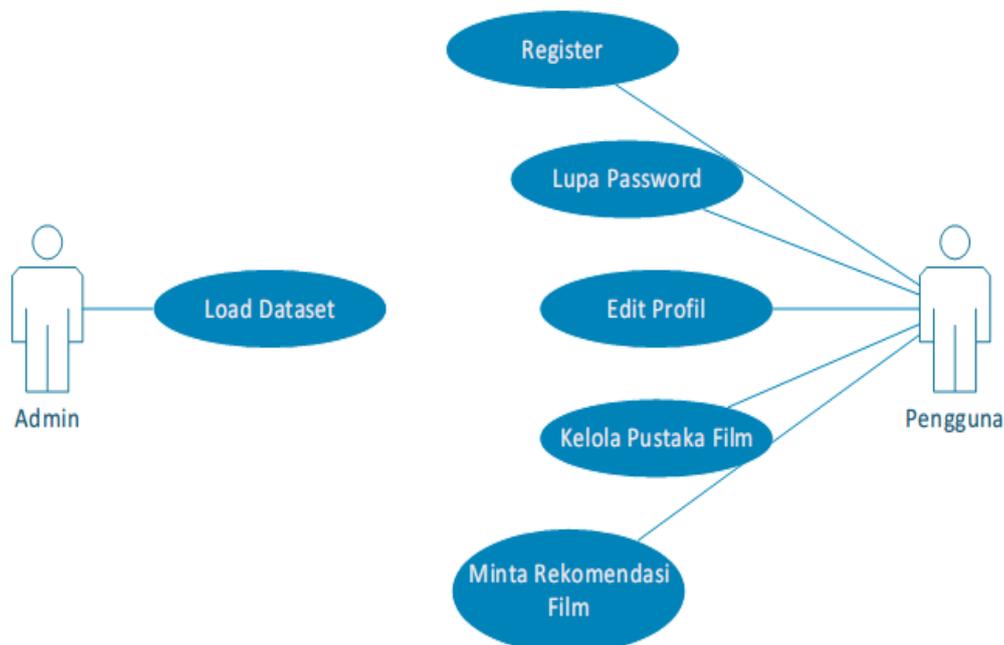
userID	movieID	rating	timestamp
296	50	5	1532993858
296	110	5	1532993855
296	296	4	1532993850
296	318	5	1532993816
296	356	5	1532993877

Berdasarkan Tabel 4 diatas dataset rating.csv merupakan dataset yang berisi seluruh rating yang telah diberikan oleh setiap user terhadap suatu film. Pada dataset ini setiap user telah memberikan rating pada sedikitnya 20 judul film. Dataset ini memiliki 100.837 baris data yang akan digunakan untuk membangun model data training dan data uji pada metode *Item Based Collaborative Filtering*.

### 2.3 Analisis dan Perancangan Sistem

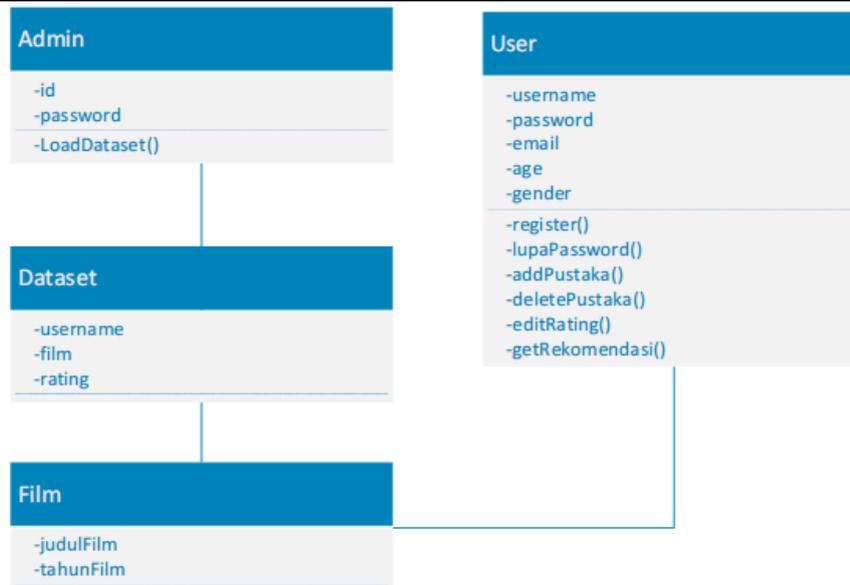
Tahap analisis dilakukan dengan 2 cara, yaitu observasi dan wawancara yang bertujuan untuk menggali semua kebutuhan yang harus dipenuhi pada sistem nantinya. Observasi dilakukan dengan mengamati sistem-sistem yang serupa, sedangkan wawancara dilakukan dengan memberikan beberapa pertanyaan kepada 10 orang dengan rentang usia dan latar belakang yang berbeda, mereka dipilih karena memiliki minat yang tinggi dalam menonton film.

Pada tahap perancangan sistem penulis menggunakan *Object Oriented Programming (OOP)* untuk menjembatani antara fase kebutuhan sistem ke fase implementasi sistemnya, sehingga mempermudah penulis dalam pembuatan kode program nantinya. Beberapa diagram digunakan dalam perancangan sistem ini yaitu meliputi *usecase diagram*, *activity diagram* dan *class diagram*. Diagram tersebut akan ditunjukkan pada Gambar 2 dan Gambar 3 seperti dibawah ini.



Gambar 2. Usecase Diagram Sistem Rekomendasi

Pada Gambar 2 ditunjukkan bahwa terdapat dua aktor yang berhubungan dengan sistem yaitu admin dan pengguna. Admin dibutuhkan oleh sistem untuk load dataset, sedangkan pengguna memerlukan beberapa fungsi yang harus ada pada sistem yaitu meliputi register, ganti password, edit profil, kelola pustaka film, dan meminta rekomendasi. Dan untuk mempresentasikan kelas-kelas yang akan digunakan pada sistem dibutuhkan class diagram seperti yang terlihat pada Gambar 3 berikut ini.



Gambar 3. Class Diagram Sistem Rekomendasi

## 2.4 Implementasi Sistem

Pada penelitian ini penulis menggunakan *Apache Mahout* sebagai framework rekomendasinya, dengan Java sebagai bahasa pemrograman serta Swing sebagai penyedia *interface* programnya. Untuk penjabaran dari metode *Item Based Collaborative Filtering* yang diterapkan pada *Apache Mahout* dapat dilihat dalam pseudocode dibawah ini.

Pseudocode :

```

public List<RecommendedItem> getUserBased(int UserID) throws
TasteException,
IOException {
String PreferencesCSVpath = System.getProperty("user.dir") +
"/Datasets/ratings.csv";
DataModel model = new FileDataModel(new File(PreferencesCSVpath));
UserSimilarity similarity = new
TanimotoCoefficientSimilarity(model);
UserNeighborhood neighborhood = new NearestNUserNeighborhood(5,
similarity, model);
Recommender recommender = new GenericUserBasedRecommender (model,
neighborhood,
similarity);
List<RecommendedItem> recommendations =
recommender.recommend(UserID, 4);
return recommendations;
}
  
```

Pada *pseudocode* diatas dijelaskan bahwa *method get Userbase* digunakan untuk mendapatkan rekomendasi film dari *Apache Mahout* dengan meminta UserID sebagai parameternya. Tahap awal pada metode *Item Based Collaborative Filtering* adalah dengan menghitung nilai kemiripan antar item yang telah dirating oleh user sebelumnya dengan menggunakan algoritma similarity. Untuk algoritma similarity yang digunakan pada penelitian ini menggunakan *Tanimoto Coeficient Similarity* dengan rumus Persamaan 1.

$$T(a, b) = \frac{N_c}{N_a + N_b - N_c} \quad (1)$$

Selanjutnya untuk menentukan neighborhood pada setiap item rekomendasi, penulis menggunakan *NearestUserNeighborhood* dengan nilai koefisien 5, hasil rekomendasi yang diminta sebagai parameter berjumlah 4, dan menampilkan hasil rekomendasi yang diminta pada GUI.

## 2.5 Pengujian

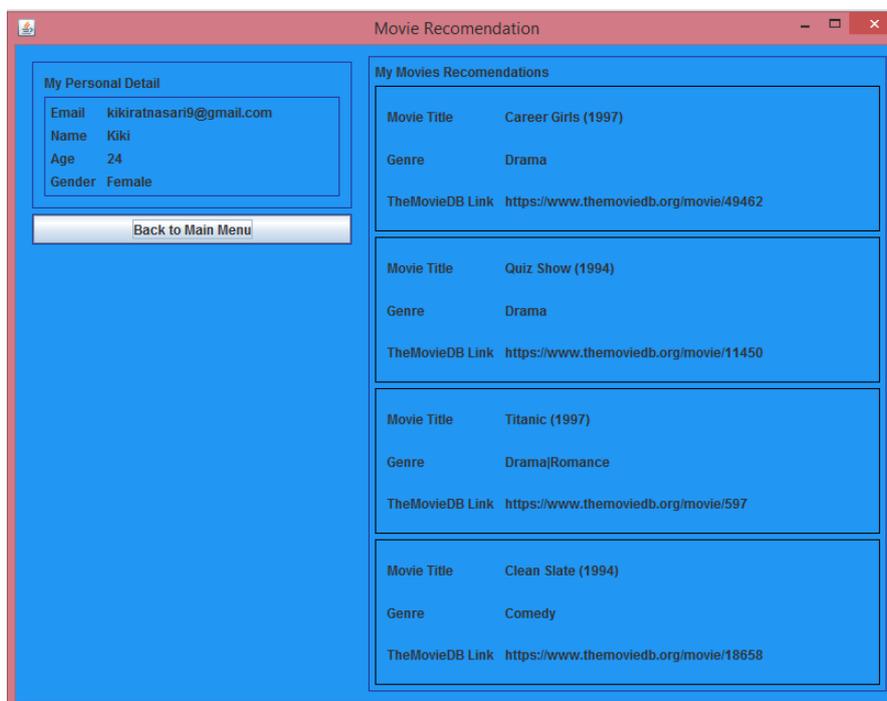
Pada penelitian kali ini pengujian dilakukan untuk mengetahui nilai keakuratan dari hasil rekomendasi yang dihasilkan oleh sistem. Untuk pengujian tersebut teknik yang dirasa sesuai dan cocok adalah cross validation, khususnya k-fold cross validation. Pengujian akan dilakukan dengan mengikuti skenario pengujian sebagai berikut :

1. Penulis mempersiapkan data sample yang berisi user serta nilai ratingnya untuk berbagai judul film.
2. Penulis menentukan nilai "K" untuk k-fold validationnya.
3. Data sample dibagi menjadi beberapa grup sesuai dengan nilai "K" yang telah ditentukan.
4. Dalam 1 grup dipilih secara acak untuk digunakan sebagai data uji, sedangkan sisanya digunakan sebagai data training.
5. Penulis menghitung nilai keakuratan dari data uji yang telah dipilih, lalu dilanjutkan dengan memilih 1 grup baru sebagai sample.
6. Setelah dilakukan beberapa kali percobaan, maka penulis akan membuat laporan hasil pengujiannya.

## 3. Hasil Penelitian dan Pembahasan

### 3.1 Hasil sistem

Setelah dilakukannya implementasi sistem dengan menggunakan metode dan framework yang dipilih pada penelitian ini menghasilkan sistem rekomendasi yang memiliki beberapa fungsi atau fitur yang dapat digunakan oleh penggunanya meliputi : register akun, login akun, edit profil, ganti password, beri rating film, hapus film, dan minta rekomendasi. Semua fungsi tersebut dapat digunakan jika pengguna telah memiliki akses masuk ke dalam sistem dengan melakukan register terlebih dahulu. Untuk mendapatkan rekomendasi user atau pengguna diminta untuk memberikan rating terlebih dahulu terhadap beberapa film yang telah mereka tonton sebelumnya, hasil dari pemberian rating tersebutlah yang akan dijadikan sebagai preferensi untuk sistem menampilkan hasil rekomendasi. Tampilan dari salah satu fitur akan ditunjukkan pada Gambar 4 yaitu tampilan dari fungsi minta rekomendasi seperti dibawah ini.



Gambar 4. Tampilan Halaman Minta Rekomendasi

### 3.2 Hasil Pengujian

Pengujian dilakukan dengan tujuan untuk mengetahui nilai keakuratan pada hasil rekomendasi yang diberikan oleh sistem setelah diterapkannya metode Item based Collaborative Filtering pada Apache Mahout. Pemilihan dataset yang digunakan pada pengujian dengan nilai

K sebesar 5 menghasilkan 5 grup dataset dengan masing-masing grup berjumlah 20167 data. 4 grup data tersebut nantinya digunakan sebagai data training dan 2 grup data digunakan sebagai data uji. Metode pengukuran nilai akurasi dan presisi yang digunakan yaitu Root Mean Square Evaluation yang telah tersedia pada Apache Mahout.

Dengan mengikuti langkah-langkah pada skenario pengujian yang telah dibuat sebelumnya, maka didapatkan hasil seperti pengujian sebagai berikut:

Evaluation Type: Root Mean Square Evaluation Similarity Type: Tanimoto Coefficient Neighborhood Type: Nearest N Neighbors For value= 5.0, evaluation score is <b>0.9727565650127541</b> Precision at 5: 0.08862745098039225 Recall at 5: 0.096111111111111104
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Terlihat dari hasil pengujian bahwa nilai akurasinya sudah sangat memuaskan yaitu sebesar 97%. Hal ini menunjukkan bahwa sistem sudah sesuai dengan preferensi dari setiap penggunanya. Nilai evaluation score pada Apache Mahout dihitung dengan membandingkan seberapa relevan hasil rekomendasi yang dihasilkan dari data training dengan data uji yang digunakan dalam model pengujian. Sedangkan nilai precision dan recall yang masih rendah menunjukkan bahwa dataset yang digunakan masih perlu ditingkatkan lagi.

Precision disini merupakan berapa banyak film dari hasil rekomendasi yang diberikan oleh sistem yang sudah benar-benar relevan terhadap penggunanya. Misalkan sistem memberikan 5 rekomendasi film, dari 5 hasil rekomendasi tersebut yang benar-benar relevan dengan pengguna hanya 3, jadi nilai precisionnya adalah 3/5.

Recall disini merupakan banyaknya film relevan yang diberikan oleh sistem dari keseluruhan film relevan yang sebenarnya. Misalnya sistem memberikan 5 rekomendasi film, dari 5 rekomendasi tersebut yang benar-benar relevan dengan pengguna adalah 3, sedangkan seharusnya ada 10 film yang relevan, jadi nilai recallnya adalah 3/10.

#### 4. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan maka dapat disimpulkan sebagai berikut :

1. Apache Mahout dapat digunakan sebagai pustaka tambahan untuk membuat model rekomendasi film dengan menggunakan metode Item Based Collaborative Filtering. Model yang dihasilkan dapat diterapkan menggunakan bahasa pemrograman Java yang menjadi penghubung antara pengguna dan sistem dengan memanfaatkan fitur GUI pada Java.
2. Penggunaan Item Based Collaborative Filtering sebagai metode rekomendasi menghasilkan nilai akurasi yang sangat memuaskan dengan besar 97%. Hal ini menunjukkan bahwa metode ini merupakan metode yang sangat tepat untuk kasus rekomendasi film. Akan tetapi pada pengembangannya masih memiliki kekurangan yaitu dalam pemrosesan data berukuran besar masih memakan waktu yang cukup lama.

#### Daftar Notasi

Na = user yang memilih item a,

Nb = user yang memilih item b,

Nc = user yang memilih item keduanya.

#### Referensi

- [1] M. K. P. Hidayat, "Model Sistem Informasi Toko Film Digital Dengan Recommender System," p. 3, 2013.
- [2] L. Dzumiroh and R. Saptono, "Penerapan Metode Collaborative Filtering Menggunakan Rating Implisit pada Sistem Rekomendasi Pemilihan Film di Rental VCD," *J. Teknol. Inf. ITSmart*, vol. 1, no. 2, p. 54, 2016.
- [3] E. W. Sumarlin, S. Hansun, and Y. W. Wiratama, "Rancang Bangun Aplikasi Rekomendasi Film dengan," vol. 10, no. 2, pp. 1244–1250, 2016.
- [4] S. Sari and D. T. Hendra, "Aplikasi Rekomendasi Film menggunakan Pendekatan Collaborative Filtering dan Euclidean Distance sebagai ukuran kemiripan rating ISBN : 979-26-0280-1 ISBN : 979-26-0280-1," pp. 135–140, 2015.

- 
- [5] G. I. Marthasari, Y. Azhar, and D. K. Puspitaningrum, "Perlengkapan Pesta Menggunakan Collaborative Filtering Dan Penggalan," vol. 5, no. 1, pp. 1–8, 2015.
  - [6] B. Utomo and Y. Suhari, "Rekomendasi Film Berbasis Web Pada Bioskop Mini Menggunakan Algoritma Nearest-Neighbor," vol. 5, no. 1, 2013.
  - [7] E. A. Laksana, "Collaborative Filtering dan Aplikasinya," *J. Ilm. Teknol. Inf. Terap.*, vol. 1, no. 1, pp. 36–40, 2014.
  - [8] Y. A. Pratama, D. Wijaya, and A. Halim, "Digital Cakery Dengan Algoritma Collaborative Filtering," vol. 14, no. 1, pp. 79–88, 2013.
  - [9] - Aminudin and M. Alwi, "Analisa Multithreading Pada Sistem Rekomendasi Menggunakan Metode Collaborative Filtering Dengan Apache Mahout," *Techno.Com*, vol. 17, no. 1, pp. 1–11, 2017.
  - [10] A. Jabakji and H. Dag, "Improving item-based recommendation accuracy with user's preferences on Apache Mahout," *Proc. - 2016 IEEE Int. Conf. Big Data, Big Data 2016*, no. 978, pp. 1742–1749, 2016.