

Implementasi Metode Collaborative Filtering Based Untuk Sistem Rekomendasi Buku Fiksi

Pharadya Ajeng Swari Sukmawati¹⁾ Lely Hiryanto²⁾ Viny Christanti Mawardi³⁾

^{1) 2) 3)} Teknik Informatika, FTI, Universitas Tarumanagara

Jl. Letjen S Parman no 1, Jakarta 11440 Indonesia

¹⁾email : pharadya.535170065@stu.untar.ac.id, ²⁾email : lelyh@fti.untar.ac.id, ³⁾email : viny@fti.untar.ac.id

ABSTRACT

A book is the result of someone's work in the form of a collection of papers containing articles intended for publication. One of the benefits of books is that they can open people's horizons and can educate reason, thoughts and faith. Indonesia is ranked 60 out of 61 countries with reading interest problems causing low public interest in reading. One of the factors that can be taken into consideration for interest in reading in Indonesia is the difficulty in finding books to read because the many kinds of books in circulation make it difficult for readers to decide which books to read, besides that readers only want to read books with the best reputation. The purpose of creating a book recommendation system is to make it easier to find fiction-type books to read. The data used in this design are book data and rating data from Kaggle. This design uses one of the recommendation system methods, namely collaborative filtering. Collaborative filtering is a recommendation method that calculates similarity between items by users to make choices. The system will recommend 5 books according to the book title that the user will input.

Key words

Book, Collaborative Filtering, Recommendation System

1. Pendahuluan

Buku merupakan sebuah hasil karya dari seseorang berupa kumpulan kertas yang tertulis atau menjadi tulisan berjilid yang ditujukan untuk sebuah penerbitan. Salah satu manfaat dari buku adalah membuka wawasan masyarakat serta dapat mencerdaskan akal, pikiran dan iman. Buku dianggap berhasil jika isi dari buku tersebut sukses dalam menambah atau menggugah minat dari pembaca dalam memahami isi dari buku tersebut. Buku yang terdapat di Indonesia memiliki dua jenis yaitu buku fiksi dan buku non fiksi. Perbedaan dari buku fiksi dan non fiksi adalah buku fiksi yaitu buku yang menggunakan bahasa kiasan dan tidak bermakna sebenarnya dengan tujuan mengajak pembaca seolah-olah masuk ke dalam cerita yang dibuat penulis, sedangkan buku non fiksi yaitu buku yang dibuat berdasarkan kisah nyata ataupun disusun berdasarkan fakta.

Menurut data *Most Littered Nation In The World* yang dilakukan oleh Central Connecticut State University (CCSU) pada Maret 2016, Indonesia menempati peringkat ke-60 dari 61 negara dengan persoalan minat baca. Dalam studi tersebut, Indonesia berada di bawah Thailand dan berada di atas Botswana. Kriteria tersebut diukur dalam studi salah satunya adalah jumlah orang dalam berkunjung ke perpustakaan untuk sekedar membaca buku. Faktanya berkunjung ke perpustakaan bukanlah budaya yang ada di Indonesia tetapi lebih suka berkunjung ke toko buku untuk membeli buku bacaan yang kita mau baca.[1]

Terdapat banyak toko buku yang ada di Indonesia, salah satunya adalah toko buku Gramedia. Toko buku tersebut memiliki toko fisik dan toko non fisik yang berupa website dan juga menjual buku diberbagai macam *e-commerce* yang ada di Indonesia. Dalam website toko Gramedia hanya dapat melihat detail buku namun tidak ada rating pada tampilan website tersebut, tetapi pengguna dapat mengetahui jika ada buku-buku yang *best seller*. Selain toko buku Gramedia terdapat website yang berisi buku - buku lokal maupun internasional yang sudah direview dan memiliki penilaian berupa rating dari penggunanya, website tersebut bernama Goodreads. Dalam website tersebut terdapat detail buku seperti judul buku, pengarang, penerbit, ISBN, tahun terbit serta sinopsis tetapi website tersebut tidak bisa membeli buku langsung di dalam website tersebut. Jika penggunanya ingin membeli buku akan diarahkan ke *marketplace* yang menjual buku tersebut seperti Amazon.

Salah satu faktor yang bisa dijadikan bahan pertimbangan terhadap minat baca yang ada di Indonesia adalah kesulitannya mencari buku yang akan dibaca karena banyaknya macam buku yang beredar mengakibatkan para pembaca mengalami kesulitan menentukan buku yang akan dibaca, selain itu pembaca hanya ingin membaca buku dengan reputasi yang terbaik, karena semakin rendah reputasinya bertanda buku tersebut sedikit para pembacanya.

Dari permasalahan yang sudah dijelaskan, maka dibuatlah sebuah sistem rekomendasi yang diharapkan dapat membantu dalam mencari buku yang diinginkan. Sistem rekomendasi tersebut akan berbasis website dengan metode collaborative filtering yang merupakan salah satu metode dari sistem rekomendasi yang menghitung nilai kemiripan antar item oleh pengguna untuk menentukan pilihan terbaik.

2. Dasar Teori

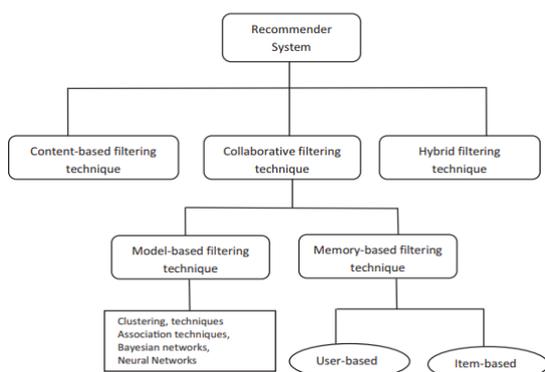
2.1 Buku

Buku adalah kumpulan atau lembaran kertas yang tertulis dan mengandung tulisan. Kumpulan kertas tersebut dijilid menjadi satu dan berisi tulisan, gambar ataupun tempelan.[2] Terdapat beberapa nama lain dari buku, salah satunya adalah kitab. Kitab berasal dari kata serapan bahasa Arab yang memiliki arti “buku” dalam bahasa Indonesia. Kitab juga merujuk pada sebuah teks atau tulisan yang dijilid menjadi satu dan merujuk kepada jenis tulisan kuno yang memiliki ketetapan hukum atau undang-undang yang mengatur.[3]

Buku terdapat dua jenis, yaitu buku fiksi dan non fiksi. Buku fiksi merupakan buku yang berisikan cerita yang tidak nyata, khayalan atau pun rekaan. Ciri khas dalam buku fiksi adalah cerita yang terkandung di dalamnya bukan merupakan kisah nyata atau hanyalah karangan fiktif belaka, sedangkan buku nonfiksi merupakan buku yang berisi karangan yang berdasarkan kejadian nyata atau disusun berdasarkan fakta yang ada. Ciri dari buku non fiksi adalah ditulis berdasarkan fakta yang ada atau bermakna sebenarnya. Contoh dari buku fiksi yaitu novel, cerita anak dan komik, sedangkan contoh buku non fiksi yaitu karya ilmiah, artikel, tesis, dll.[4]

2.2 Sistem Rekomendasi

Sistem rekomendasi adalah sistem oenyaringan informasi yang menangani masalah kelebihan informasi dengan memfilter fragmen informasi penting dari sejumlah besar informasi yang dihasilkan secara dinamis sesuai dengan preferensi pengguna, minat atau perilaku yang diamati tentang item[5]. Sistem rekomendasi memiliki kemampuan untuk memprediksi apakah pengguna lebih memilih item tertentu atau tidak berdasarkan profil penggunaannya[5].



Gambar 1 Teknik Sistem Rekomendasi [5]

Dalam penggunaan sistem rekomendasi terdapat berbagai macam teknik yang efisien dan akurat agar memberikan sistem rekomendasi yang baik dan berguna untuk setiap pengguna. Hal ini menjelaskan pentingnya

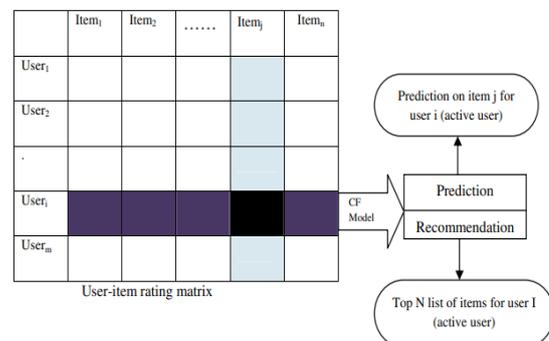
memahami fitur dan potensi dari teknik sistem rekomendasi yang berbeda. Pada **Gambar 1** menampilkan macam-macam teknik dari rekomendasi yang berbeda. Teknik tersebut diantaranya adalah[5] :

1. Content Based Filtering

Content Based Filtering adalah metode yang bergantung pada domain dan lebih menekankan pada analisis atribut item untuk menghasilkan prediksi. Di dalam content based filtering, rekomendasi dibuat berdasarkan profil pengguna menggunakan fitur yang diambil dari konten item yang telah dievaluasi pengguna di masa lalu. Item yang sebagian besar terkait dengan nilai positif direkomendasikan kepada pengguna[5].

2. Collaborative Filtering

Collaborative filtering adalah teknik prediksi domain independen untuk konten yang tidak dapat dengan mudah dan memadai dijelaskan oleh metadata seperti film dan musik. Teknik atau metode collaborative filtering bekerja dengan membangun basis data preferensi untuk item oleh pengguna. Yang kemudian mencocokkan pengguna dengan minat dan preferensi yang relevan dengan perhitungan kesamaan antar profil untuk membuat rekomendasi.[6] Pengguna membangun grup yang disebut lingkungan. Seorang pengguna mendapatkan rekomendasi dari item yang belum diberi rating sebelumnya.



Gambar 2 Proses Collaborative Filtering [5]

Rekomendasi yang dihasilkan oleh collaborative filtering dapat berupa prediksi ataupun rekomendasi. Prediksi adalah nilai numerik R_{ij} menyatakan prediksi skor item j untuk untuk pengguna I , sedangkan rekomendasi adalah daftar item N teratas yang paling disukai pengguna seperti yang ditunjukkan pada **Gambar 2**. Teknik collaborative filtering dapat dibagi menjadi dua kategori yaitu *memory based* dan *model based*.

3. Hybrid Filtering

Hybrid filtering adalah teknik rekomendasi yang menggabungkan teknik yang berbeda untuk mendapatkan optimasi sistem yang lebih baik untuk menghindari keterbatasan dan masalah sistem

rekomendasi murni.[5] Ide dari teknik hybrid ini adalah mengkombinasikan metode agar memberikan hasil yang lebih akurat dan rekomendasi efektif dari metode tunggal sebagai kerugian satu metode yang dapat diatasi oleh metode lain[5]. Menggunakan beberapa teknik rekomendasi dapat menekan kelemahan teknik individu dalam model gabungan.

2.3. Collaborative Filtering

Collaborative Filtering merupakan proses dari penyaringan atau evaluasi item menggunakan pendapat orang lain. Collaborative filtering berakar dari sesuatu yang pernah dilakukan manusia selama berabad-abad dengan berbagi pendapat orang lain.[7]

Metode collaborative filtering dapat dibagi menjadi dua kategori, yaitu *memory based* dan *model based*. *Memory based* membutuhkan rating, item dan pengguna yang disimpan ke dalam memori, sedangkan *model based* merupakan metode yang berbasis model yang dibuat secara berkala dengan membuat ringkasan pola peringkat secara *offline*. [8]

Metode collaborative filtering *memory based* dapat dilakukan dengan dua cara yaitu dengan *user based* dan *item based*. *User based* menghitung kesamaan antara pengguna dengan membandingkan peringkat mereka pada item yang sama, dan kemudian menghitung peringkat yang diprediksi untuk suatu item dengan pengguna aktif sebagai rata – rata tertimbang dari peringkat item oleh pengguna yang mirip dengan pengguna aktif. Sedangkan *item based* adalah menghitung prediksi menggunakan kesamaan antar item dan bukan kesamaan antar pengguna. prediksi dibuat dengan mengambil rata – rata tertimbang dari peringkat pengguna aktif pada item serupa.[5]

Proses pembuatan model dalam *model based* dapat dilakukan menggunakan teknik *machine learning* atau teknik data mining. Dalam penggunaan pembelajaran metode juga telah mengubah cara rekomendasi dari merekomendasikan apa yang harus dikonsumsi oleh pengguna untuk merekomendasikan kapan harus benar – benar mengkonsumsi suatu produk. Berikut adalah metode pembelajaran yang digunakan oleh sistem rekomendasi *model based* [5] :

1. Association Rule

Metode *Association Rule* mengekstrak aturan yang memprediksi terjadinya suatu item berdasarkan pada kehadiran item lain dalam transaksi.

2. Clustering

Metode *clustering* mencoba mempartisi sekumpulan data menjadi sebuah set sub-cluster untuk menemukan kelompok yang bermakna yang ada dalam diri mereka. Metode *clustering* yang baik akan menghasilkan cluster yang berkualitas dimana kesamaan intra-cluster tinggi, sedangkan kesamaan antar-cluster rendah. K-Means dan *Self-Organizing*

Map (SOM) adalah yang paling umum digunakan dalam *clustering*. *Clustering* dapat digunakan untuk mengurangi himpunan kandidat dalam metode collaborative filtering.

3. Decision Tree

Decision Tree adalah metode yang didasarkan pada metodologi grafik pohon yang dibangun dengan menganalisis serangkaian contoh pelatihan label kelasnya diketahui. *Decision tree* juga juga fleksibel dalam menangani item dengan campuran fitur bernilai nyata dan kategoris juga sebagai item yang memiliki beberapa fitur khusus yang hilang.

4. Artificial Neural Network

ANN adalah struktur dari banyak neuron yang terhubung yang disusun berlapis – lapis secara sistematis. ANN memiliki kemampuan memperkirakan fungsi nonlinier dan menangkap kompleks hubungan dalam kumpulan data agar bisa efisien dan merata beroperasi jika bagian dari jaringan gagal.

5. Regression

Regression digunakan ketika dua atau lebih variable dianggap terhubung secara sistematis oleh sebuah hubungan linier.

2.4. Item-Based Nearest Neighbor

Item-based nearest neighbor merupakan transpose dari salah satu metode collaborative filtering yaitu user-based nearest neighbor. Pada user-based metode menghasilkan rekomendasi berdasarkan kesamaan antar pengguna sedangkan item-based menghasilkan rekomendasi berdasarkan kesamaan antar item.[9]

Metode Pearson Correlation Coefficient merupakan metode yang umum digunakan untuk menghitung persamaan antar objek yang didasarkan oleh perhitungan korelasi linier antar dua himpunan yang berbeda.[10] Berikut ini adalah persamaan metode pearson correlation[10]:

$$sim(i,j) = corr_{i,i} \frac{\sum_{j=1}^k (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{i=1}^k (R_{u,i} - \bar{R}_i)^2 \sum_{j=1}^k (R_{u,j} - \bar{R}_j)^2}} \quad (1)$$

Dimana :

$Sim(i,j)$ = kemiripan item i dan item j

$R_{u,i}$ adalah rating yang pernah diberikan pengguna u terhadap item i

\bar{R}_i adalah rata-rata rating dari item i

$R_{u,j}$ adalah rating yang pernah diberikan pengguna u terhadap item j

\bar{R}_j adalah rata-rata rating dari item j

Perhitungan prediksi digunakan untuk memprediksi nilai rating yang diberikan pengguna terhadap objek atau item tertentu. Dengan metode *neighborhood-based*, rating yang diberikan oleh pengguna langsung tersimpan di sistem dan kemudian digunakan untuk memprediksi rating item baru. Adapun bentuk persamaan algoritma K-NN adalah sebagai berikut[10] :

$$P_{u,i} = \frac{\sum_{j=1}^k (sim(i, j) \times R_{u,j})}{\sum_{j=1}^k (sim(i, j))} \quad (2)$$

Dimana :

$P_{u,i}$ adalah prediksi rating untuk user u

$sim(i, j)$ adalah kemiripan item k dan item i

$R_{u,j}$ adalah rating diberikan pengguna u terhadap item j

2.5. Mean Absolute Error

Mean Absolute Error (MAE) merupakan metode yang digunakan untuk menghitung tingkat akurasi atau nilai error untuk menghitung prediksi rating dari sistem terhadap rating yang sebenarnya diberikan pengguna terhadap suatu sistem.[10] Berikut ini merupakan persamaan dari Mean Absolute Error (MAE) [10]:

$$MAE_i = \frac{1}{n} \sum_{i \in E} (|p_i - r_i|) \quad (3)$$

Dimana:

MAE_i adalah model evaluasi algoritma

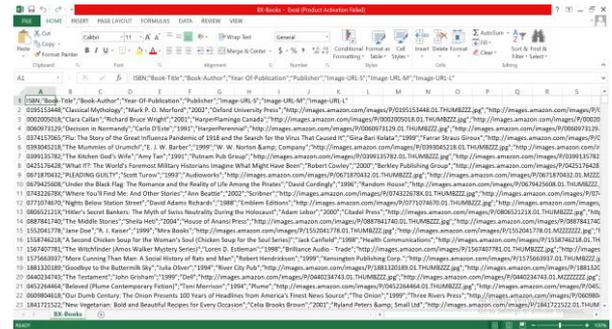
r_i adalah nilai rating yang sebenarnya

p_i adalah nilai rating prediksi

n adalah banyaknya pasangan rating asli dan rating prediksi

3. Hasil Percobaan

Hasil percobaan perhitungan terhadap metode Collaborative Filtering berjenis Item-Based menggunakan data rating buku yang sudah dikumpulkan. Data tersebut merupakan data buku, data rating buku dan data user. Data tersebut bersumber dari website kaggle. Percobaan ini menggunakan 706 data buku yang sudah memiliki rating dari 806 user. Data tersebut terkumpul dalam format csv yang dapat dilihat pada **Gambar 3**, terdapat 3 data yang terkumpul yaitu data buku, data rating dan data user. Data buku terdiri dari judul buku ISBN, penerbit, penulis, tahun terbit dan URL gambar. Data rating terdiri dari ISBN, User-ID dan rating. Data user terdiri dari user-id, lokasi dan umur.



Gambar 3 Contoh Data yang Dipakai

Langkah pertama adalah menggabungkan dua data yaitu data buku dengan data rating menjadi satu, kemudian data tersebut mencari rata – rata dari masing – masing rating buku yang sudah terkumpul. Data tersebut diurutkan mana yang rating tertinggi dan terbanyak kemudian disusun menjadi matriks.

Langkah selanjutnya adalah menghitung cosine similarity dari buku dengan judul tertentu. Mencari nilai cosine similarity bertujuan mencari buku yang memiliki nilai kesamaan antar item. Pada tahap ini akan dicari 5 rekomendasi buku yang sesuai dengan pencarian judul buku yang diinput. Pada **Tabel 2** hasil rekomendasi buku berdasarkan kesamaan antar item yang dihitung dengan cosine similarity.

Tabel 1 Hasil Cosine Similarity Buku 1984

Judul Buku	Nilai cosine similarity
Animal Farm	0,27
The Handmaid’s Tale	0,26
Brave New World	0,23
The Vampire Lestat (Vampire Chronicles, Book II)	0,23
The Hours : A Novel	0,22

Tabel 2 Hasil Cosine Similarity Buku Animal Farm

Judul Buku	Nilai Cosine Similarity
1984	0,27
Angus, Thongs and Full-Frontal Snogging: Confessions of Georgi a Nicolson	0,24
Midnight	0,23
Second Nature	0,21
Call of The Wild	0,20

Dari hasil cosine similarity didapatkan buku yang direkomendasikan dengan judul 1984 yaitu Animal Farm, The Handmaid’s Tale, Brave New World, The Vampire Lestat (Vampire Chronicles, Book II) dan The Hours : A Novel. Sedangkan rekomendasi buku dengan judul Animal Farm adalah 1984, Angus, Thongs and Full-Frontal Snogging: Confessions of Gergia Nicolson, Mignight, Second Nature dan Call of The Wild.

Perhitungan akurasi dilakukan menggunakan metode Mean Absolute Error (MAE) yang digunakan

untuk menjadi kesalahan atau nilai error dari metode yang dipakai. Error tersebut dapat dihitung dengan cara mencari selisih dari nilai rating yang sebenarnya dengan nilai rating prediksi. Nilai MAE dari sistem rekomendasi menggunakan metode collaborative filtering adalah 0,83.

4. Kesimpulan

Dari hasil data pengujian yang telah dilakukan sebelumnya, dapat ditarik beberapa kesimpulan diantaranya sebagai berikut:

1. Sistem rekomendasi menggunakan metode collaborative filtering dengan jenis item-based dapat memberikan rekomendasi buku kepada pengguna.
2. Untuk menentukan nilai kemiripan antar buku pada metode collaborative filtering menggunakan cosine similarity.
3. Dalam sistem rekomendasi tidak ditemukannya kesalahan dalam menentukan rekomendasi buku.

Adapun saran yang diberikan diantaranya adalah metode ini dapat digabungkan dengan jenis metode collaborative filtering yang lainnya.

REFERENSI

- [1] Muharram, Riza Miftah. Minat Baca Orang Indonesia Rendah, Masa Sih?. <https://www.ruangguru.com/blog/benarkah-minat-baca-orang-indonesia-rendah>. 22 Februari 2023.
- [2] Merriam. Definition of Book. www.merriam-webster.com. 1 Maret 2023.
- [3] Pratiwi, Inten Esti. Sejarah Lahirnya Buku, Bermula Dari Coretan Gambar di Dinding Gua. <https://www.kompas.com/tren/read/2021/05/16/190000165/sejarah-lahirnya-buku-bermula-dari-coretan-gambar-di-dinding-gua?page=all#:~:text=Dikutip%20dari%20berbagai%20sumber%20sejarah,rakyat%20Mesir%20menciptakan%20kertas%20papyrus.&text=Namun%20ada%20pula%20yang%20mengatakan,Awalnya%20media%20buku%20sangat%20beragam>, 1 Maret 2023.
- [4] Bisma, Leo. Perbedaan Buku Fiksi & Nonfiksi: Pengertian, Ciri, Jenis, Contoh | Bahasa Indonesia Kelas 12. <https://www.ruangguru.com/blog/perbedaan-buku-fiksi-dan-nonfiksi>. 1 Maret 2023.
- [5] F.O. Isinkaye, Y.O. Foljimi, B.A. Ojokoh, 2015, "Recommendation System: Principles, methods and Evaluation", *Egyptian Informatics Journal*
- [6] Herlocker JL, et.al, "Evaluating Collaborative Filtering Recommender System", *ACM Trans Inform Syst*, Vol. 1, 2004 Schafer, J.Ben; et al., 2007. "Collaborative Filtering Recommender Systems". *The Adaptive Web: Methods and Strategies of Web Personalization*, Vol.4321.
- [7] Breese, J.S., Heckerman, D., Kadie, C, 1998, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", In *Proceeding of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, Morgan Kaufmann, Madison.
- [8] Sarwar, B., Karypis, G., Konstan, J.A., Riedl, J, 2001, "Item—Based Collaborative Filtering Recommendation Algorithms", *Proceedings of the 10th international*

conference on World Wide Web, ACM Press, Hong Kong.

- [9] Lubis, Yohana Imelda, Napitupulu, Dohar Josua, Dharma, Arie Satria, Implementasi Metode Hybrid Filtering (Collaborative dan Content-Based) Untuk Sitem Rekomendasi Pariwisata, <https://citee.ft.ugm.ac.id/download51.php?f=TI-5%20-%20Implementasi%20Metode%20Hybrid%20Filtering.pdf>, 13 Maret 2023.

Pharadya Ajeng Swari Sukmawati, Seorang Mahasiswi program studi Teknik Informatika Universitas Tarumanagara, Jakarta.

Lely Hiryanto, Memperoleh gelar S.Kom dari Universitas Tarumanagara tahun 2001. Kemudian memperoleh gelar M.Sc. dan Ph.D. dari Curtin University of Technology tahun 2006 dan 2022. Saat ini aktif sebagai Dosen Tetap Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta.

Viny Christanti Mawardi, Memperoleh gelar S.Kom. dari Universitas Tarumanagara tahun 2004. Kemudian memperoleh gelar M.Kom. dari Universitas Indonesia tahun 2008. Saat ini aktif sebagai Dosen Tetap Ffakultas Teknologi Informasi Universitas Tarumanagara, Jakarta.