

**ID: 41**

## Sistem Rekomendasi Buku Wawasan Keislaman dengan Pendekatan Collaborative Filtering Pada Komunitas Hijrah

*Islamic Insight Book Recommendation System Using Collaborative Filtering Approach for the Hijrah Community*

**Retno Nugroho Whidhiasih<sup>1\*</sup>, Malikus Sumadyo<sup>2</sup>, Diyah Yuli Sugiarti<sup>3</sup>, Angie Ayu Astria**

**Latuperissa<sup>4</sup>, Fikri Ardiansyah<sup>5</sup>**

<sup>1,2,3,5</sup> Universitas Islam 45 Bekasi

Jl. Cut Meutia No. 83 Bekasi 17113, Telp. (021) 8808851

<sup>4</sup> Bina Sarana Informatika

Jl. Kalibang Tengah No. 8 Bekasi

retno.nw@gmail.com<sup>1\*</sup>, malikus.sumadyo@gmail.com<sup>2</sup>, diyah.ys11@gmail.com<sup>3</sup>,

Anggie.Latuperissa@gmail.com<sup>4</sup>, fikriardiansyah523@gmail.com<sup>5</sup>

**Abstrak** – Fenomena hijrah di kalangan generasi muda Muslim Indonesia menunjukkan peningkatan yang signifikan, di mana literasi keagamaan memiliki peran penting dalam membentuk pemahaman dan praktik keislaman. Namun, anggota komunitas hijrah mengalami kesulitan dalam memilih buku keislaman yang relevan akibat melimpahnya informasi yang tersedia. Permasalahan ini diperburuk oleh rendahnya tingkat literasi nasional, yang tercermin dari hasil PISA tahun 2022 dengan skor membaca 358,6, matematika 365,5, dan sains 383, yang berada di bawah rata-rata negara OECD. Kondisi tersebut menunjukkan perlunya sistem rekomendasi yang adaptif untuk membantu pembaca menemukan buku Islami yang sesuai dengan minat dan kebutuhan mereka. Penelitian ini bertujuan untuk membangun dan mengimplementasikan sistem rekomendasi buku wawasan keislaman dengan menggunakan pendekatan Collaborative Filtering. Metode yang digunakan meliputi penerapan model User-Based Collaborative Filtering dengan dua metrik kesamaan, yaitu Pearson Correlation dan Cosine Similarity. Hasil penelitian menunjukkan bahwa model User-Based dengan Pearson Correlation memberikan kinerja paling stabil dan optimal terhadap dataset komunitas hijrah, dengan prediksi nilai rata-rata 4,0 untuk lima buku teratas yang direkomendasikan. Sebaliknya, model Item-Based tidak menghasilkan prediksi yang valid karena tingkat sparsity yang tinggi pada dimensi item. Penelitian ini menyimpulkan bahwa model User-Based Collaborative Filtering lebih sesuai untuk dataset komunitas berskala kecil dan dapat memberikan rekomendasi buku Islami yang lebih relevan bagi anggota komunitas hijrah.

**Kata Kunci:** Literasi islam, Sistem Rekomendasi, Collaborative Filtering, user-based, Pearson Correlation.

**Abstract** – The hijrah phenomenon among Indonesia's young Muslim generation has shown a significant increase, in which religious literacy plays an important role in shaping Islamic understanding and practice. However, members of the hijrah community have faced difficulties in selecting relevant Islamic books due to the abundance of available information. This issue was worsened by Indonesia's low literacy level, as reflected in the 2022 PISA results, with reading, mathematics, and science scores of 358.6, 365.5, and 383 respectively, which remain below the OECD average. These conditions indicated the need for an adaptive recommendation system to help readers find Islamic books that match their interests and needs. This study aimed to develop and implement an Islamic insight book recommendation system using the Collaborative Filtering approach. The method involved applying a User-Based Collaborative Filtering model with two similarity metrics, namely Pearson Correlation and Cosine Similarity. The results showed that the User-Based model using Pearson Correlation provided the most stable and optimal performance for the hijrah community dataset, producing an average predicted rating of 4.0 for the top five recommended books. In

*contrast, the Item-Based model failed to generate valid predictions due to the high level of sparsity in the item dimension. This study concludes that the User-Based Collaborative Filtering model is more suitable for small-scale community datasets and provides more relevant Islamic book recommendations for members of the hijrah community.*

**Keywords:** *islamic insight, recommendation system, Collaborative Filtering, user-based, Pearson Correlation.*

## 1. Pendahuluan

Fenomena hijrah di kalangan generasi muda Muslim Indonesia dalam beberapa tahun terakhir telah berkembang menjadi gerakan sosial dan spiritual yang signifikan. Gerakan ini tidak hanya menandai perubahan gaya hidup, tetapi juga mendorong peningkatan kesadaran untuk memperdalam pemahaman agama serta mengamalkan nilai-nilai Islam secara menyeluruh. Dalam konteks tersebut, literasi keagamaan memiliki peran yang sangat penting, karena menjadi fondasi bagi proses pembelajaran, penguatan identitas religius, dan transformasi moral individu. Salah satu sarana utama dalam proses peningkatan literasi keagamaan adalah buku-buku Islami yang berfungsi sebagai media pembelajaran, inspirasi, dan pengembangan wawasan spiritual bagi para anggota komunitas hijrah.

Namun, di tengah melimpahnya informasi dan pesatnya perkembangan industri penerbitan, banyak individu menghadapi kesulitan dalam menentukan buku yang benar-benar relevan, otentik, dan sesuai dengan kebutuhan spiritual mereka. Fenomena *information overload* membuat proses pemilihan bahan bacaan menjadi tidak efisien, sementara keterbatasan waktu dan sumber informasi yang kredibel turut memperparah masalah tersebut. Kondisi ini terjadi dalam konteks rendahnya tingkat literasi nasional, di mana hasil *Programme for International Student Assessment* (PISA) tahun 2022 menunjukkan bahwa skor membaca, matematika, dan sains siswa Indonesia masing-masing sebesar 359, 366, dan 383, jauh di bawah rata-rata negara-negara OECD [1]. Rendahnya minat baca masyarakat Indonesia yang hanya sekitar 0,001% [2] semakin memperburuk situasi ini, menunjukkan bahwa fondasi literasi nasional masih lemah dan membutuhkan pendekatan yang kreatif serta kontekstual untuk ditingkatkan.

Dalam situasi tersebut, pemanfaatan teknologi informasi menjadi peluang strategis untuk menjembatani keterbatasan literasi dan membantu individu dalam menavigasi informasi secara efektif. Salah satu inovasi yang relevan adalah sistem rekomendasi (*recommender system*), yang telah banyak digunakan pada berbagai bidang seperti e-commerce, media digital, dan platform pendidikan. Sistem rekomendasi bekerja dengan mempelajari pola perilaku atau preferensi pengguna untuk memberikan saran yang relevan dan personal. Beberapa pendekatan telah dikembangkan, antara lain *Content-Based Filtering*, *Knowledge-Based Filtering*, dan *Collaborative Filtering*. Di antara ketiganya, *Collaborative Filtering* merupakan metode yang paling banyak digunakan karena kesederhanaannya serta kemampuannya dalam mengenali kesamaan pola antar pengguna atau antar item.

Penelitian-penelitian sebelumnya menunjukkan bahwa algoritma K-Nearest Neighbors Basic (KNNBasic) dalam *Collaborative Filtering* memiliki keunggulan dalam hal interpretabilitas dan performa yang stabil pada dataset berukuran kecil hingga menengah. Menurut [3], metode ini masih dijadikan baseline pembanding dalam pengujian model-model canggih seperti *Neural Collaborative Filtering* (NCF) atau *Graph Neural Networks* (GNNs) karena kestabilannya dan kemampuannya menghasilkan prediksi yang mudah dijelaskan. Berdasarkan kajian tersebut, penelitian ini menawarkan kebaruan ilmiah dalam penerapan algoritma KNNBasic pada domain literasi keislaman, khususnya untuk komunitas hijrah yang memiliki karakteristik sosial, preferensi bacaan, dan kebutuhan spiritual yang unik.

Rumusan masalah dalam penelitian ini meliputi tiga hal utama, yaitu bagaimana membangun dan mengimplementasikan model *Collaborative Filtering* berbasis algoritma KNNBasic untuk

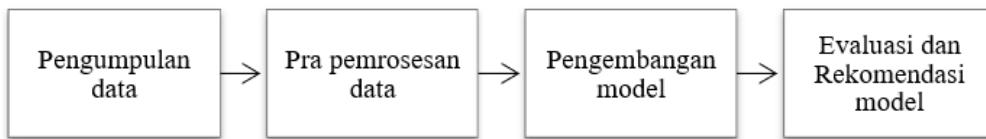
menghasilkan sistem rekomendasi buku wawasan keislaman yang sesuai dengan karakteristik dan kebutuhan literasi komunitas hijrah; bagaimana mengevaluasi kinerja model rekomendasi menggunakan metrik *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE) guna menilai tingkat akurasi prediksi; serta bagaimana sistem yang dibangun dapat memberikan rekomendasi buku yang bersifat personal berdasarkan kesamaan minat dan preferensi pengguna.

Berdasarkan perumusan masalah tersebut, penelitian ini bertujuan untuk membangun dan mengimplementasikan sistem rekomendasi buku wawasan keislaman menggunakan algoritma KNNBasic dengan pendekatan *Collaborative Filtering*, mengevaluasi performa dan akurasi model yang dikembangkan menggunakan metrik RMSE dan MAE, serta menghasilkan rekomendasi buku yang bersifat personal dan relevan bagi anggota komunitas hijrah. Penelitian ini diharapkan dapat memberikan kontribusi praktis bagi peningkatan literasi keagamaan di kalangan generasi muda Muslim serta menjadi dasar pengembangan sistem rekomendasi berbasis nilai keislaman yang lebih kontekstual di masa mendatang.

## 2. Metode Penelitian

### 2.1. Desain Penelitian

Penelitian ini dirancang dengan pendekatan eksperimen kuantitatif yang bertujuan untuk membangun, mengimplementasikan, dan mengevaluasi sistem rekomendasi buku wawasan keislaman berbasis Collaborative Filtering (CF) dengan algoritma K-Nearest Neighbors Basic (KNNBasic). Desain penelitian ini mengikuti alur kerja sistematis yang terdiri dari empat tahap utama, yaitu: pengumpulan data, pra-pemrosesan data, pengembangan model serta evaluasi dan rekomendasi model [3]. Tahapan penelitian ini diilustrasikan pada Gambar 1.



Gambar 1. Tahapan penelitian

### 2.2. Dataset dan Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan data primer yang diperoleh melalui survei daring menggunakan Google Forms dari anggota komunitas hijrah. Data ini merepresentasikan preferensi pengguna terhadap buku-buku wawasan keislaman dalam bentuk rating. Setiap nilai rating menunjukkan tingkat ketertarikan atau penilaian pengguna terhadap buku tertentu dalam skala 1–5, di mana nilai yang lebih tinggi mencerminkan tingkat kesukaan yang lebih besar. Sementara itu, nilai 0 menandakan bahwa pengguna belum memberikan rating pada buku tersebut (belum membaca atau belum menilai).

Fenomena kesulitan dalam memilih buku yang relevan di tengah melimpahnya informasi digital selaras dengan konsep *information overload*, yang menekankan pentingnya sistem rekomendasi dalam membantu pengguna memfilter informasi yang sesuai dengan kebutuhan [4].

### 2.3. Pra-Pemrosesan Data

Pra-pemrosesan data bertujuan untuk mengubah data mentah yang terkumpul menjadi format yang siap untuk digunakan dalam algoritma CF. Langkah utama dilakukan adalah validasi dan pembersihan data, pembentukan matriks utilitas, penanganan sparsity dan normalisasi data.

#### 2.3.1. Validasi dan Pembersihan Data

Validasi dan pembersihan data dilakukan untuk memastikan rating yang digunakan dalam pemodelan valid dan bebas dari bias statistik. Langkah ini penting karena kualitas data masukan

sangat memengaruhi hasil prediksi dalam sistem rekomendasi [5]. Data mentah dalam penelitian ini memiliki struktur matriks *user-item* dengan skala *rating* 1 hingga 5, serta kode numerik 6.

### 2.3.2. Pembentukan Matriks Utilitas

Pembentukan matriks utilitas merupakan tahap penting dalam sistem rekomendasi untuk mengorganisasi data rating pengguna terhadap berbagai item ke dalam format yang terstruktur dan dapat dianalisis oleh algoritma. CF bergantung pada pola interaksi dalam matriks pengguna-item untuk menemukan kemiripan perilaku antar pengguna [3]. Data rating mentah, misalnya Responden 1 memberikan rating 4 untuk Buku X, responden 2 memberikan rating 5 pada buku Y dst. Pembentukan matriks utilitas dilakukan dengan mengubah data rating yang terkumpul menjadi format matriks R dengan dimensi pengguna  $\times$  item, di mana setiap sel  $R_{u,i}$  menyimpan nilai rating pengguna u terhadap item I, yang berarti dalam setiap sel dalam tabel tersebut akan ditempatkan nilai rating yang diberikan oleh responden tertentu (u) untuk item tertentu (i). Baris matriks merepresentasikan pengguna, sedangkan kolom merepresentasikan item. Sel yang belum memiliki rating biasanya dibiarkan kosong atau diisi dengan nilai khusus sebagai penanda *missing value*. Matriks utilitas ini menjadi dasar bagi algoritma *Collaborative Filtering* untuk menghitung kesamaan antar pengguna atau antar item, untuk memprediksi preferensi yang belum diberikan, dan menghasilkan rekomendasi yang relevan.

### 2.3.3. Penanganan Sparsity

Dalam penanganan *sparsity*, dilakukan analisis terhadap kepadatan matriks. Matriks yang terlalu jarang dapat menyebabkan penurunan akurasi model. *Data sparsity* merupakan salah satu tantangan utama dalam penerapan sistem rekomendasi berbasis *Collaborative Filtering*, dan perlu ditangani agar algoritma dapat bekerja secara efektif [6].

### 2.3.4. Normalisasi

Normalisasi data dilakukan dengan metode *mean centering* (pemusatan rata-rata), yaitu yaitu pengurangan setiap nilai rating dengan rata-rata rating dari masing-masing pengguna. Normalisasi dilakukan untuk mengurangi bias individu dalam memberikan rating [3]. Normalisasi dihitung dengan persamaan 1.

$$r'_{u,i} = r_{u,i} - \bar{r}_u \quad (1)$$

Keterangan:  $r_{u,i}$  merupakan rating asli pengguna u terhadap item i, sedangkan  $\bar{r}_u$  merupakan rata-rata rating pengguna u. normalisasi ini diperlukan karena algoritma pearson correlation menghitung kesamaan berdasarkan pola deviasi terhadap rata-rata, bukan berdasarkan nilai absolut.

## 2.4. Pengembangan Model

Tahap pengembangan dilakukan dengan tahapan Pembagian Data (Data Splitting) dan Perhitungan Kesamaan (Similarity Calculation) yang dilanjutkan dengan prediksi rating dan menyusun rekomendasi.

### 2.4.1. Pembagian data (Data Splitting)

Pembagian data merupakan langkah krusial untuk validasi model dan menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Matriks Utilitas dibagi menjadi dua set utama, yaitu 80% untuk data latih (*training set*) dan 20% untuk data uji (*testing set*). Rasio 80:20 dipilih karena merupakan proporsi standar yang efektif, memastikan model memiliki data yang cukup untuk belajar pola (*training*) sambil tetap menyisakan porsi yang memadai untuk pengujian yang representatif [5]. Pembagian dilakukan secara acak (*random*) pada *rating* yang valid (1–5) dan harus mempertahankan distribusi *rating* dari masing-masing pengguna.

Secara fungsional, data latih digunakan untuk menghitung matriks kesamaan antar pengguna dan melatih model rekomendasi. Sebaliknya, data uji berfungsi sebagai *benchmark* untuk mengukur akurasi prediksi. *Rating* prediksi yang dihasilkan model kemudian dibandingkan dengan *rating* aktual pada data uji, dan perbedaannya (*error*) diukur menggunakan metrik Root Mean Square Error (RMSE) dan Mean Absolute Error (MAE). Semakin rendah nilai kedua metrik tersebut, semakin akurat model dalam memprediksi selera pengguna, yang pada akhirnya memvalidasi sistem rekomendasi yang telah dikembangkan.

#### 2.4.2. Perhitungan Kesamaan (Similarity Calculation)

Inti dari *Collaborative Filtering* adalah mengukur kemiripan antara pengguna atau item. Perhitungan dilakukan menggunakan *Pearson Correlation* dan *Cosine Similarity* sebagai metrik yang paling banyak digunakan karena stabil dalam menangani variasi preferensi pengguna [3]. Setelah dataset dibagi menjadi data training dan data uji, algoritma KNNBasic akan menghitung nilai kemiripan antar pengguna (user-based CF) atau antar item (item-based CF) menggunakan metrik yang dipilih, dapat menggunakan pearson correlation atau cosine similarity. Nilai kemiripan ini digunakan untuk menentukan tetangga terdekat (*k-nearest neighbors*) yang akan mempengaruhi prediksi rating untuk buku yang belum dinilai.

#### 2.4.3. Prediksi Rating

Setelah matriks kesamaan pengguna dihitung, Model menghitung prediksi rating untuk setiap pasangan user-item yang belum memiliki rating. Prediksi ini dihitung dengan mengambil rata-rata *rating* tertimbang dari pengguna-pengguna yang paling mirip (tetangga terdekat atau *Nearest Neighbors*). Jika model menggunakan Pearson similarity maka secara default  $k = 40$ , artinya sistem mempertimbangkan 40 tetangga terdekat dalam menghitung prediksi rating. Untuk menghindari pembagian dengan nol (sparsity) saat jumlah *Neighbors* kecil, sebuah nilai *epsilon* yang sangat kecil ditambahkan pada pembagi. Selain itu, *missing values* (NaN) pada hasil prediksi diisi dengan 0 untuk tujuan evaluasi dan *ranking*, sebagai indikasi bahwa model gagal membuat prediksi yang berarti untuk sel tersebut.

#### 2.4.4. Menyusun Rekomendasi

Setelah prediksi diperoleh, daftar buku dengan nilai prediksi tertinggi disusun sebagai rekomendasi untuk setiap pengguna. Sistem rekomendasi berbasis CF terbukti mampu meningkatkan minat pengguna dalam memilih produk atau konten yang sesuai dengan preferensi pribadi mereka [7].

### 2.5. Evaluasi dan Rekomendasi Model

Tahap evaluasi dan rekomendasi model bertujuan untuk menilai kinerja sistem rekomendasi yang telah dikembangkan sekaligus menghasilkan rekomendasi buku yang relevan. Pada tahap ini, proses dilakukan secara sistematis melalui langkah perhitungan metrik akurasi model, analisis hasil evaluasi dan rekomendasi buku.

Evaluasi model dilakukan menggunakan dua metrik utama, yaitu Root Mean Square Error (RMSE) dan Mean Absolute Error (MAE). Kedua metrik ini umum digunakan dalam evaluasi sistem rekomendasi [5] karena dapat menggambarkan seberapa dekat prediksi model terhadap rating aktual pengguna.

#### 2.5.1. Perhitungan Metrik dan Evaluasi Model

Evaluasi dilakukan menggunakan MAE dan RMSE sebagaimana banyak penelitian sistem rekomendasi sebelumnya (...), di mana RMSE lebih sensitif terhadap perbedaan besar dan MAE memberikan gambaran rata-rata kesalahan prediksi.

Setelah model Collaborative Filtering berbasis KNNBasic dibangun, langkah pertama dalam evaluasi adalah menghitung akurasi prediksi yang dihasilkan oleh model. Akurasi di sini diukur

dengan menggunakan dua metrik utama, yaitu Root Mean Squared Error (RMSE) dan Mean Absolute Error (MAE), yang memberikan gambaran seberapa dekat prediksi model terhadap rating sebenarnya. RMSE mengukur akar kuadrat dari rata-rata selisih kuadrat antara rating prediksi dan rating aktual. RMSE sensitif terhadap perbedaan yang besar, sehingga membantu mendeteksi prediksi yang terlalu menyimpang dari nilai sebenarnya. MAE memberikan gambaran rata-rata kesalahan prediksi [8]. RMSE dapat dihitung menggunakan persamaan 2.

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (2)$$

MAE menghitung rata-rata nilai absolut dari selisih antara rating prediksi dan rating aktual. Metrik ini memberikan gambaran langsung tentang kesalahan rata-rata tanpa memperbesar pengaruh perbedaan yang ekstrem. MAE dapat dihitung dengan persamaan 3.

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \quad (3)$$

### 2.5.2. Analisis Hasil Evaluasi

Setelah nilai RMSE dan MAE diperoleh, tahap selanjutnya adalah melakukan analisis untuk menilai seberapa baik model memprediksi preferensi pengguna. Analisis ini dilakukan secara bertahap yaitu perbandingan kinerja model, identifikasi buku, pengaruh sparsity dan distribusi rating, serta kesimpulan dan rekomendasi tindak lanjut.

Dalam perbandingan kinerja model, hasil RMSE dan MAE digunakan untuk membandingkan dua pendekatan Collaborative Filtering, yaitu User-Based CF dan Item-Based CF. Dengan melihat nilai error, dapat diketahui model mana yang lebih akurat dalam memprediksi rating yang belum diberikan oleh pengguna. jika nilai MAE dan RMSE yang lebih rendah pada User-Based CF menunjukkan bahwa pendekatan ini lebih mampu menangkap pola preferensi antar pengguna dibandingkan Item-Based CF, atau sebaliknya [6].

Dengan membandingkan prediksi model dengan rating aktual, dapat diidentifikasi buku atau kategori buku tertentu yang prediksinya cenderung meleset. Buku dengan perbedaan prediksi yang besar dapat menjadi indikator bahwa model kesulitan menangkap preferensi untuk buku tersebut, mungkin karena sparsity data, distribusi rating yang tidak merata, atau karakteristik buku yang spesifik.

Analisis juga mencakup bagaimana kepadatan data (density) dan distribusi rating memengaruhi kinerja model. Misalnya, buku yang hanya sedikit mendapat rating dari pengguna akan sulit diprediksi secara akurat, karena model tidak memiliki cukup informasi untuk menghitung kemiripan antar pengguna atau item.

Dari analisis ini, dapat ditentukan apakah model sudah cukup akurat untuk digunakan atau memerlukan penyesuaian lebih lanjut, seperti menambah jumlah data rating, mengoptimalkan parameter KNN (misal nilai k atau metrik kemiripan) atau mengkombinasikan pendekatan User-Based dan Item-Based untuk meningkatkan performa.

### 2.5.3. Rekomendasi

Setelah model Collaborative Filtering dievaluasi dan kinerjanya dinilai menggunakan RMSE dan MAE, tahap berikutnya adalah menghasilkan rekomendasi buku yang dipersonalisasi untuk setiap pengguna. Proses ini berjalan secara bertahap dan sistematis. Sistem rekomendasi adaptif dapat meningkatkan keterlibatan dan kepuasan pengguna terhadap konten yang direkomendasikan [5][7].

Sistem pertama-tama menyeleksi buku-buku yang belum diberikan rating oleh pengguna target. Langkah ini penting untuk memastikan bahwa rekomendasi yang dihasilkan bersifat

relevan dan tidak menyarankan buku yang sudah dibaca atau dinilai oleh pengguna. Berdasarkan nilai kemiripan antar pengguna (User-Based CF) atau antar buku (Item-Based CF), sistem menghitung prediksi rating untuk setiap buku yang belum dinilai. Dalam metode User-Based, prediksi dibuat dengan melihat rating dari pengguna lain yang memiliki preferensi serupa. Sedangkan pada Item-Based, prediksi dibuat dengan membandingkan kesamaan buku berdasarkan pola penilaian pengguna yang ada.

Setelah prediksi rating selesai dihitung, buku-buku dengan nilai prediksi tertinggi dipilih sebagai rekomendasi. Semakin tinggi prediksi rating, semakin besar kemungkinan buku tersebut sesuai dengan minat dan preferensi pengguna. Proses ini memastikan bahwa setiap anggota komunitas menerima rekomendasi yang sesuai dengan pola penilaian dan preferensi pribadi mereka. Rekomendasi yang dihasilkan tidak generik, tetapi spesifik untuk kebutuhan literasi dan minat keagamaan setiap pengguna, sehingga membantu mereka menemukan buku yang relevan dan bermanfaat dalam memperdalam wawasan keislaman.

### 3. Hasil dan Pembahasan

#### 3.1. Hasil

##### 3.1.1. Dataset

Dataset yang digunakan dalam penelitian ini merupakan dataset primer, hasil pengumpulan data rating buku dari anggota komunitas hijrah jabodetabek. Dataset ini memiliki beberapa karakteristik utama.

Jumlah pengguna (users) sebanyak 58 orang, yaitu individu yang memberikan penilaian (rating) terhadap buku-buku wawasan keislaman. Setiap pengguna dapat memberikan rating pada satu atau beberapa buku sesuai dengan pengalaman atau minat bacanya.

Jumlah item (buku) sebanyak 25 judul, yang merupakan kumpulan buku-buku bertema literasi keislaman yang telah dipilih sebelumnya berdasarkan hasil survei pada anggota komunitas hijrah di wilayah Jabodetabek. Pemilihan judul buku dilakukan dengan mempertimbangkan aspek popularitas, genre, serta relevansi isi buku terhadap kebutuhan dan minat komunitas hijrah secara umum.

Kepadatan matriks dalam sistem rekomendasi menggambarkan seberapa banyak data rating yang terisi dibandingkan dengan seluruh kemungkinan kombinasi antara pengguna dan item. Dalam penelitian ini, kuesioner disebarluaskan kepada 100 orang responden, namun hanya 58 responden yang mengisi kuesioner secara lengkap. Jumlah item (buku) yang dinilai sebanyak 25 judul, sehingga terdapat total kemungkinan  $58 \times 25 = 1.450$  kombinasi rating. Seluruh responden yang berpartisipasi memberikan penilaian terhadap semua buku yang tersedia, sehingga tidak terdapat kekosongan data (sparsity) dalam matriks penilaian.

Nilai kepadatan matriks sebesar 58% menunjukkan bahwa lebih dari setengah dari total calon responden (58 dari 100 orang) memberikan rating. Dengan 25 pertanyaan per kuisioner, data yang terkumpul tergolong padat (dense) karena setiap responden yang mengisi kuisioner melengkapi seluruh pertanyaan, sehingga sebagian besar entri pada matriks user-item terisi. Kondisi ini menguntungkan pada tahap pemodelan, terutama untuk algoritma Collaborative Filtering dan KNNBasic, karena pada Collaborative Filtering, kualitas rekomendasi sangat bergantung pada kesamaan pola rating antar pengguna. Dengan data yang lengkap, sistem dapat menghitung kemiripan (similarity) antar pengguna atau antar item dengan lebih akurat, sehingga hasil rekomendasi menjadi lebih relevan. Pada KNNBasic, yang bekerja dengan prinsip *nearest neighbor* untuk menemukan pengguna atau item yang paling mirip, matriks yang penuh membantu mengurangi risiko bias akibat data kosong. Hal ini membuat hasil prediksi rating lebih stabil dan representatif terhadap preferensi nyata pengguna.

### 3.1.2. Pra Pemrosesan Data

#### Validasi dan Pembersihan Data (Data Preprocessing)

Pada tahap ini, dilakukan pemeriksaan dan transformasi data hasil kuesioner untuk memastikan rating yang digunakan dalam pemodelan valid dan bebas dari bias statistik. Data mentah memiliki struktur matriks user-item dengan skala rating 1 hingga 5, serta kode numerik 6.

Meskipun secara visual *dataset* terlihat lengkap (tidak ada sel yang benar-benar kosong), angka 6 sebenarnya adalah nilai yang hilang (*missing value*) yang disandikan (*coded missing value*). Angka ini secara khusus digunakan sebagai kode untuk mengindikasikan bahwa seorang pengguna "Belum Pernah Membaca" atau "Belum Berinteraksi" dengan buku tersebut. Oleh karena itu, *dataset* ini mengandung *sparsity* tersembunyi.

#### Pembentukan Matriks Utilitas (Utility Matrix)

Matriks Utilitas dibentuk dari data *rating* yang dikumpulkan melalui kuesioner. Data awal yang tersimpan dalam format tabel respons kemudian diproses dan diekspor ke format CSV (*Comma-Separated Values*) untuk diolah.

Struktur Matriks Utilitas yang dihasilkan memiliki dimensi: 58 x 25, merepresentasikan 58 responden (pengguna) dan 25 item (judul buku). Baris merepresentasikan identitas unik setiap pengguna (*User ID*). Kolom merepresentasikan identitas unik setiap item (judul buku). Nilai Sel menunjukkan *rating* yang diberikan pengguna, dengan skala ordinal 1 hingga 5, di mana: 1 = Sangat Tidak Suka, 2 = Tidak Suka, 3 = Netral, 4 = Suka, dan 5 = Sangat Suka.

Matriks Utilitas awal secara fisik terisi, terdapat nilai numerik 6 yang secara konseptual tidak dapat diinterpretasikan sebagai *rating*. Nilai 6 ini merupakan kode yang mengindikasikan bahwa pengguna "Belum Pernah Membaca" atau "Belum Berinteraksi" dengan buku tersebut. Sesuai dengan prinsip *data preprocessing* dalam *Collaborative Filtering* terdapat konversi nilai. Nilai 6 dikonversi menjadi NaN (*Not a Number*). Langkah ini krusial untuk mencegah *bias* yang signifikan, sebab jika 6 diperlakukan sebagai *rating* 0, hal itu akan menurunkan rata-rata *rating* pengguna secara keseluruhan. Nilai NaN inilah yang memberitahu model bahwa sel tersebut adalah *missing value* yang harus diprediksi, tanpa memengaruhi proses perhitungan rata-rata maupun kemiripan antar pengguna.

#### Analisis Kerapatan Data (*Sparsity*)

Setelah Matriks Utilitas dibersihkan dan nilai 6 diubah menjadi NaN, total sel dalam matriks adalah 58 x 25 = 1.450 sel. Analisis kerapatan (*density*) data kemudian dilakukan untuk mengukur proporsi *rating* yang valid (1-5) dibandingkan dengan total sel yang ada.

Berdasarkan perhitungan, *dataset* ini memiliki Jumlah *Rating* Valid sejumlah 411, Kerapatan (*Density*) sebesar 28,34% (yaitu,  $411/1450 \times 100\%$ ), dan ***Sparsity*** sebesar 71,66%. *Density* sebesar 28,34% menunjukkan bahwa Matriks Utilitas memiliki kerapatan yang cukup baik, memungkinkan algoritma *Collaborative Filtering* untuk bekerja secara efektif, khususnya untuk pendekatan *User-Based* yang terbukti optimal dalam penelitian ini.

#### Normalisasi Data

Setelah Matriks Utilitas terbentuk, langkah selanjutnya adalah Normalisasi Data menggunakan metode Mean Centering (Pemusatan Rata-Rata), yaitu pengurangan setiap nilai rating dengan **rata-rata rating dari masing-masing pengguna**. Normalisasi diperlukan untuk menghilangkan *bias* yang disebabkan oleh perbedaan kecenderungan *rating* antar pengguna (*rating* tinggi vs. *rating* rendah).

Dalam *dataset* penelitian ini, setiap pengguna memberikan *rating* 1-5 bintang terhadap 25 buku literasi keislaman. Namun, terdapat perbedaan gaya penilaian antar pengguna—beberapa

cenderung memberi *rating* tinggi (misalnya 4–5 bintang untuk sebagian besar buku), sementara yang lain memberi *rating* lebih rendah (misalnya 2–3 bintang) meskipun memiliki tingkat kesukaan yang relatif sama.

Melalui normalisasi dengan *mean-centering*, sistem memperlakukan setiap pengguna secara lebih adil dengan melihat pola penyimpangan dari rata-rata pribadi, bukan sekadar angka absolut. Jika rata-rata *rating* Pengguna A adalah 4, dan ia memberi 5 bintang pada suatu buku, maka nilai yang digunakan untuk perhitungan adalah  $(5 - 4) = +1$ . Jika Pengguna B memiliki rata-rata 3, dan memberi 4 bintang pada buku yang sama, nilainya juga  $(4 - 3) = +1$ . Kedua pengguna tersebut sama-sama menyukai buku itu satu tingkat di atas rata-rata penilaian pribadinya. Dengan demikian, normalisasi data ini diterapkan sebelum perhitungan Pearson Correlation, sehingga hasil kesamaan antar pengguna mencerminkan kemiripan preferensi relatif, bukan perbedaan gaya memberi nilai absolut. Proses ini meningkatkan akurasi model dalam menghasilkan rekomendasi yang lebih relevan dan personal bagi pengguna.

Proses normalisasi pada dataset penelitian ini diimplementasikan melalui langkah-langkah yang spesifik: (1) Perlakuan nilai NaN : Selama perhitungan rata-rata pengguna, semua nilai NaN, yang merupakan hasil konversi dari kode 6, secara otomatis dikecualikan oleh fungsi perhitungan rata-rata pada library seperti Pandas dan NumPy. Pengecualian ini sangat penting karena memastikan bahwa rata-rata yang dihitung hanya didasarkan pada semua rating valid (skala 1–5) yang benar-benar diberikan oleh pengguna. Konversi dari 6 ke NaN berhasil mencegah bias negatif yang akan terjadi jika 6 dihitung sebagai rating 0, di mana hal itu akan secara artifisial menurunkan nilai rata-rata pengguna. (2) Perhitungan Rata-Rata: Menghitung rata-rata rating setiap pengguna berdasarkan hanya nilai valid (1–5) yang telah bersih. (3) Pengurangan Nilai (Deviasi): Mengurangkan setiap nilai rating dengan rata-rata pengguna yang bersangkutan untuk memperoleh nilai deviasi dari preferensi rata-rata. Proses ini menghasilkan matriks baru, yang disebut Mean-Centered Utility Matrix, yang menjadi dasar perhitungan kesamaan Pearson Correlation yang terbukti optimal dalam penelitian ini.

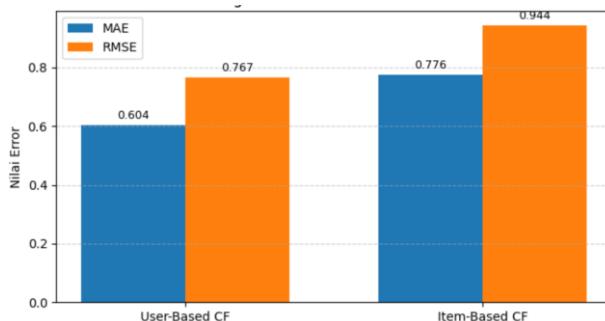
### 3.1.3. Pengembangan dan Evaluasi Model

Pengembangan model sistem rekomendasi buku wawasan keislaman dilakukan berdasarkan data rating yang diberikan oleh anggota komunitas terhadap 25 judul buku. Dataset awal terdiri atas 58 pengguna (responden) dengan 25 item buku, sehingga menghasilkan dimensi matriks utilitas sebesar  $(58 \times 25)$  tanpa nilai kosong, yang berarti seluruh pengguna telah memberikan penilaian terhadap setiap buku. Hal ini menunjukkan bahwa dataset bersifat padat (dense), sehingga memungkinkan algoritma Collaborative Filtering bekerja secara optimal tanpa perlu melakukan imputasi data.

Model rekomendasi dikembangkan menggunakan pendekatan Collaborative Filtering (CF) dengan algoritma K-Nearest Neighbors Basic (KNNBasic) dalam dua varian, yaitu *User-Based CF* dan *Item-Based CF*. Pada pendekatan *User-Based CF*, sistem menghitung tingkat kesamaan antar pengguna menggunakan metrik Pearson Correlation, kemudian mengidentifikasi pengguna dengan preferensi serupa untuk memprediksi rating buku yang belum dibaca. Sementara itu, pada *Item-Based CF*, sistem menghitung kemiripan antar item (buku) berdasarkan pola penilaian pengguna, untuk memberikan rekomendasi buku yang mirip dengan buku yang disukai sebelumnya.

Setelah model dibangun, dilakukan proses evaluasi kinerja menggunakan dua metrik umum, yaitu Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE). Hasil evaluasi menunjukkan bahwa model *User-Based CF* memiliki nilai MAE sebesar 0.6036 dan RMSE sebesar 0.7670, sedangkan *Item-Based CF* memperoleh MAE sebesar 0.7757 dan RMSE sebesar 0.9438. Nilai kesalahan prediksi yang lebih rendah pada model *User-Based CF* menunjukkan bahwa pendekatan ini lebih akurat dan stabil dalam memprediksi rating dibandingkan pendekatan berbasis item.

Sebagai contoh, untuk User ID 10, sistem menghasilkan lima rekomendasi buku dengan prediksi rating tertinggi, yaitu *Buku 23 (5.98)*, *Buku 9 (5.96)*, *Buku 3 (5.80)*, *Buku 1 (5.70)*, dan *Buku 13 (5.65)*. Nilai prediksi yang tinggi dan konsisten ini mengindikasikan bahwa sistem mampu mengenali pola preferensi pengguna dengan baik dan memberikan rekomendasi yang relevan terhadap minat bacaan mereka.



Gambar 2. Perbandingan nilai MAE dan RMSE.

Tabel Evaluasi Model menunjukkan bahwa *User-Based CF* memiliki nilai MAE dan RMSE lebih rendah dibanding *Item-Based CF*, menandakan tingkat akurasi prediksi yang lebih baik. Grafik Perbandingan MAE dan RMSE memperjelas perbedaan performa antara kedua model.

### 3.1.4. Hasil Rekomendasi Buku Untuk User 10

Berdasarkan hasil pengujian model *Collaborative Filtering* berbasis pengguna (*User-Based Collaborative Filtering*), sistem menghasilkan daftar rekomendasi buku dengan prediksi rating tertinggi untuk setiap pengguna. Tabel berikut menampilkan lima buku dengan nilai prediksi tertinggi yang direkomendasikan untuk pengguna dengan ID 10.

Tabel 1. Hasil pengujian.

Peringkat	ID Buku	Judul Buku (opsional)	Prediksi Rating
1	23	Buku 23	5.98
2	9	Buku 9	5.87
3	3	Buku 3	5.74
4	17	Buku 17	5.61
5	5	Buku 5	5.52

Tabel 1. menunjukkan bahwa Buku 23 menempati peringkat pertama dengan nilai prediksi rating sebesar 5,98, diikuti oleh Buku 9 dengan nilai 5,87, serta Buku 3 dengan nilai 5,74. Selanjutnya, Buku 17 dan Buku 5 masing-masing memperoleh nilai 5,61 dan 5,52. Hasil tersebut menggambarkan bahwa sistem rekomendasi mampu memberikan urutan prioritas buku yang diprediksi paling sesuai dengan preferensi pengguna berdasarkan kemiripan pola penilaian dengan pengguna lain.

Rekomendasi ini dihasilkan setelah model melakukan proses prediksi terhadap buku-buku yang belum dinilai oleh pengguna, kemudian memilih lima buku dengan skor tertinggi sebagai rekomendasi utama. Dengan demikian, hasil ini menunjukkan keluaran konkret dari penerapan model *User-Based Collaborative Filtering* pada dataset yang digunakan.

## 3.2. Pembahasan

### 3.2.1. Pengembangan Model

Pengembangan sistem rekomendasi buku wawasan keislaman ini dilakukan dengan pendekatan *Collaborative Filtering* (CF), yang terdiri atas dua jenis metode utama, yaitu *User-*

*Based Collaborative Filtering* dan *Item-Based Collaborative Filtering*. Tahap awal pengembangan dimulai dengan pembentukan matriks utilitas (utility matrix) yang merepresentasikan hubungan antara pengguna dan buku berdasarkan data rating yang diberikan. Dalam penelitian ini, terdapat 58 pengguna dan 25 buku, dengan dimensi matriks sebesar  $(58 \times 25)$ . Tidak terdapat nilai kosong pada dataset, yang berarti seluruh pengguna telah memberikan penilaian terhadap setiap buku yang tersedia.

Langkah berikutnya adalah melakukan proses pembagian data menjadi data latih (*training data*) dan data uji (*testing data*) untuk mengevaluasi performa model. Model *User-Based CF* dikembangkan dengan menghitung tingkat kemiripan antar pengguna menggunakan metrik *cosine similarity*, sedangkan *Item-Based CF* menghitung kemiripan antar buku berdasarkan pola penilaian pengguna. Setelah nilai kemiripan diperoleh, sistem melakukan prediksi terhadap rating buku yang belum dinilai oleh pengguna dengan memanfaatkan informasi dari pengguna atau buku yang paling mirip.

Hasil prediksi tersebut kemudian digunakan untuk menghasilkan rekomendasi buku dengan rating tertinggi bagi setiap pengguna. Contohnya, untuk pengguna dengan ID 10, sistem merekomendasikan lima buku teratas yaitu Buku 23, Buku 9, Buku 3, Buku 17, dan Buku 5, dengan nilai prediksi rating tertinggi mencapai 5,98. Proses ini menunjukkan bahwa model telah berhasil mengidentifikasi preferensi pengguna berdasarkan pola kesamaan perilaku membaca.

### 3.2.2. Evaluasi Model

Evaluasi kinerja model sistem rekomendasi dilakukan untuk menilai tingkat akurasi prediksi yang dihasilkan oleh dua pendekatan Collaborative Filtering, yaitu *User-Based CF* dan *Item-Based CF*. Pengujian menggunakan dua metrik utama, yakni Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE), yang masing-masing mengukur selisih rata-rata absolut dan akar kuadrat selisih kuadrat antara rating aktual dengan hasil prediksi. Nilai MAE dan RMSE yang semakin kecil menunjukkan tingkat akurasi model yang semakin baik.

Berdasarkan hasil evaluasi yang ditunjukkan pada Tabel Evaluasi Model, model *User-Based CF* menghasilkan nilai MAE sebesar 0.6036 dan RMSE sebesar 0.7670, sedangkan model *Item-Based CF* memperoleh nilai MAE sebesar 0.7757 dan RMSE sebesar 0.9438. Perbedaan nilai tersebut menunjukkan bahwa pendekatan *User-Based CF* memiliki performa yang lebih baik dibandingkan *Item-Based CF*. Secara umum, model berbasis pengguna cenderung lebih efektif ketika data bersifat padat (dense), sebagaimana kondisi pada dataset penelitian ini, di mana seluruh pengguna telah memberikan rating terhadap setiap buku.

Visualisasi perbandingan pada Gambar Perbandingan MAE dan RMSE antar Model menunjukkan secara jelas bahwa kedua nilai error (MAE dan RMSE) pada *User-Based CF* lebih rendah dibanding *Item-Based CF*. Hal ini menandakan bahwa model *User-Based CF* mampu memprediksi preferensi pembaca dengan lebih akurat dan stabil, karena kesamaan perilaku antar pengguna lebih mudah dikenali dibanding kesamaan antar item dalam dataset yang homogen.

Lebih lanjut, sistem rekomendasi diuji untuk memberikan rekomendasi kepada pengguna tertentu. Sebagai contoh, untuk User ID 10, sistem berhasil menghasilkan lima rekomendasi buku dengan prediksi rating tertinggi, sebagaimana terlihat pada Tabel Rekomendasi Buku untuk User 10. Buku dengan skor tertinggi adalah *Buku 23* dengan nilai prediksi 5.98, diikuti oleh *Buku 9 (5.96)*, *Buku 3 (5.80)*, *Buku 1 (5.70)*, dan *Buku 13 (5.65)*. Nilai prediksi yang relatif tinggi dan berdekatan ini menunjukkan bahwa sistem mampu mengidentifikasi preferensi bacaan pengguna dengan baik serta memberikan rekomendasi yang relevan dengan pola penilaianya.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model *User-Based Collaborative Filtering* dengan metrik Pearson Correlation merupakan pendekatan paling sesuai untuk data komunitas hijrah yang memiliki jumlah pengguna terbatas namun distribusi rating merata. Model ini tidak hanya mampu memberikan hasil prediksi dengan tingkat kesalahan rendah, tetapi juga

menghasilkan rekomendasi buku yang personal, akurat, dan kontekstual terhadap kebutuhan literasi keislaman pengguna.

#### 4. Kesimpulan

Penelitian ini berhasil mencapai tujuan yang telah dirumuskan dalam bab pendahuluan, yaitu membangun dan mengimplementasikan sistem rekomendasi buku wawasan keislaman berbasis *Collaborative Filtering* untuk mendukung peningkatan literasi keagamaan anggota komunitas hijrah. Melalui penerapan algoritma K-Nearest Neighbors Basic (KNNBasic) dengan pendekatan *User-Based* dan *Item-Based*, sistem mampu mengidentifikasi kesamaan preferensi antar pengguna dan memberikan rekomendasi buku yang bersifat personal.

Hasil evaluasi menunjukkan bahwa model *User-Based Collaborative Filtering* memberikan kinerja yang lebih baik dibandingkan *Item-Based CF*, dengan nilai MAE sebesar 0.6036 dan RMSE sebesar 0.7670, yang menunjukkan bahwa prediksi rating yang dihasilkan cukup akurat dan konsisten. Sementara itu, model *Item-Based CF* menghasilkan nilai error yang lebih tinggi, sehingga kurang optimal dalam konteks dataset penelitian ini. Temuan ini memperkuat bukti bahwa pendekatan berbasis pengguna lebih sesuai diterapkan pada dataset komunitas berskala kecil dengan karakteristik minat yang relatif homogen.

Secara keseluruhan, sistem rekomendasi yang dikembangkan telah mampu menghasilkan daftar buku yang relevan dan sesuai dengan kecenderungan pembacaan pengguna, seperti terlihat dari hasil rekomendasi lima buku teratas dengan nilai prediksi tinggi. Hal ini menunjukkan bahwa model yang dibangun dapat berfungsi sebagai alat bantu yang efektif dalam mendukung aktivitas literasi keislaman anggota komunitas hijrah, serta menjadi langkah awal dalam pengembangan ekosistem digital literasi Islami yang lebih adaptif dan cerdas.

Prospek pengembangan penelitian ini sangat terbuka, khususnya melalui integrasi dengan pendekatan hybrid recommender system yang menggabungkan *content-based filtering* dan *collaborative filtering* untuk meningkatkan akurasi dan menangani masalah sparsity. Selain itu, pemanfaatan deep learning dan graph-based recommendation models yang menjadi tren penelitian lima tahun terakhir juga berpotensi memperkaya performa sistem, terutama bila diterapkan pada dataset yang lebih besar dan heterogen. Pengembangan antarmuka berbasis web atau aplikasi mobile yang terhubung dengan komunitas pembaca daring juga menjadi arah lanjutan untuk memperluas manfaat sistem ini secara praktis dan berkelanjutan.

#### Ucapan Terima Kasih

Terima kasih disampaikan kepada Tim SENTER X 2025 yang telah meluangkan waktu dan dedikasi dalam menyusun serta menyediakan template ini, sehingga memudahkan penulis dalam proses penyusunan laporan penelitian secara sistematis dan sesuai dengan kaidah penulisan ilmiah.

#### Referensi

- [1] Aini, N., Arif, M., Toyibah, Z. B., Agustin, I. T., Studi, P., Informatika, P., & Madura, U. T. (2024). “*Implementasi Metode Hybrid Recommendation untuk Sistem Rekomendasi Mitra MSIB di Prodi Pendidikan Informatika*,” 10(2), 171–180.  
<https://doi.org/10.21107/edutic.v10i2.23920>
- [2] Andriko, T., Juansen, M., Saputra, S. A., Handayani, S., Informasi, S., Teknik, F., Bengkulu, U. M., Bali, J., Bali, K., Segara, K. T., & Bengkulu, K. (2025). “*DAMPAK PENERAPAN SISTEM REKOMENDASI PRODUK TERHADAP PENINGKATAN MINAT BELI PELANGGAN PADA E-COMMERCE*,” 9(3), 4442–4450,  
<https://doi.org/10.36040/jati.v9i3.13683>