

## **HYBRID METHOD USING NON-NEGATIVE MATRIX FACTORIZATION AND KEYWORD-BASED FILTERING FOR RECOMMENDER SYSTEM IN MOOCS**

Valleryan Virgil Zuluskandar<sup>\*1</sup>, Mochammad Yusa<sup>2</sup>, Endina Putri Purwandari<sup>3</sup>

<sup>1,2</sup>Informatics, Engineering Faculty, Universitas Bengkulu, Indonesia

<sup>3</sup>Information Systems, Engineering Faculty, Universitas Bengkulu, Indonesia

Email: <sup>1</sup>[valleryan1212@gmail.com](mailto:valleryan1212@gmail.com), <sup>2</sup>[mochammad.yusa@unib.ac.id](mailto:mochammad.yusa@unib.ac.id), <sup>3</sup>[endinaputri@unib.ac.id](mailto:endinaputri@unib.ac.id)

(Article received: September 16, 2024; Revision: October 07, 2024; published: February 20, 2025)

### **Abstract**

*Massive Open Online Courses (MOOCs), introduced by Dave Cormier in 2008, have revolutionized education by providing widespread access to open and participatory online learning. While MOOCs offer broad access and flexibility in learning, users often encounter challenges in selecting appropriate courses. This leads to high dropout rates. To address this issue, this research develops a recommendation system employing the Weighted Hybrid method that combines Non-Negative Matrix Factorization (NMF) and Keyword-Based Filtering (KBF). The primary objective of the research is to enhance the accuracy of course recommendations on MOOCs. The findings of this study demonstrate that the Weighted Hybrid method, integrating NMF and KBF, successfully attained a Mean Average Precision (MAP) of 0.1963. This figure signifies an improvement compared to the MAP value of 0.1855 achieved in prior research. This method effectively addresses challenges such as cold start and sparsity, while also improving scalability. Consequently, the Weighted Hybrid approach holds promise for improving the quality of recommendations, enhancing the user's learning experience, and potentially reducing dropout rates in MOOCs.*

**Keywords:** Hybrid Method, Keyword-Based Filtering, Massive Open Online Courses, Non-Negative Matrix Factorization, Recommender System.

## **HYBRID METHOD MENGGUNAKAN NON-NEGATIVE MATRIX FACTORIZATION DAN KEYWORD-BASED FILTERING UNTUK SISTEM REKOMENDASI DALAM MOOCS**

### **Abstrak**

*Massive Open Online Courses (MOOCs) yang diperkenalkan oleh Dave Cormier pada tahun 2008 telah merevolusi pendidikan dengan memungkinkan akses luas terhadap pembelajaran online yang terbuka dan partisipatif. Meskipun MOOCs menawarkan akses luas dan fleksibilitas belajar, pengguna sering menghadapi tantangan dalam memilih kursus yang sesuai. Hal ini menyebabkan tingginya angka dropout. Untuk mengatasi masalah ini, penelitian ini mengembangkan sistem rekomendasi menggunakan metode Weighted Hybrid yang mengkombinasikan Non-Negative Matrix Factorization (NMF) dan Keyword-Based Filtering (KBF). Tujuan utama penelitian adalah untuk meningkatkan akurasi rekomendasi kursus pada MOOCs. Hasil penelitian ini menunjukkan bahwa metode Weighted Hybrid yang menggabungkan NMF dan KBF berhasil mencapai nilai Mean Average Precision (MAP) sebesar 0,1963. Angka ini menunjukkan peningkatan dibandingkan dengan nilai MAP sebesar 0,1855 yang dicapai oleh penelitian sebelumnya. Metode ini efektif dalam mengatasi tantangan seperti cold start dan sparsity, serta meningkatkan skalabilitas. Dengan demikian, pendekatan Weighted Hybrid menjanjikan untuk meningkatkan kualitas rekomendasi, meningkatkan pengalaman belajar pengguna dan potensial dalam mengurangi tingkat dropout dalam MOOCs.*

**Kata kunci:** Hybrid Method, Keyword-Based Filtering, Massive Open Online Courses, Non-Negative Matrix Factorization, Sistem Rekomendasi.

### **1. PENDAHULUAN**

Pendidikan saat ini mengalami revolusi berkat perkembangan teknologi informasi dan komunikasi dengan munculnya *Massive Open Online Courses*

(*MOOCs*). Dave Cormier pada tahun 2008 memperkenalkan istilah ini untuk pertama kali [1], yang sejak saat itu mengalami pertumbuhan peserta dan penyedia *MOOCs* secara signifikan [2]. Hingga

2021, peserta *MOOCs* mencapai 220 juta, mengalami peningkatan sebesar 40 juta pengguna baru dibandingkan tahun sebelumnya [3]. George Siemens adalah inspirator pengenalan *MOOCs* ke dalam pendidikan mendefinisikannya sebagai “kursus online dengan pendaftaran gratis dan terbuka, kurikulum terbuka, dan hasil terbuka” [4]. Sedangkan Jacobs menggambarannya sebagai simbol demokratisasi pendidikan [5]. *MOOCs* menggabungkan aset virtual terbuka yang didukung oleh para pakar di bidangnya dan mengedepankan partisipasi peserta kursus terhadap tujuan pembelajaran, pengetahuan dan kemampuan yang ada [6].

*MOOCs* (*Massive Open Online Courses*) menawarkan mata pelajaran secara gratis dan memperkenalkan strategi pengajaran revolusioner seperti pembelajaran informal, yang sesuai dengan preferensi belajar siswa [7]. Selama pandemi COVID-19, *MOOCs* menjadi alat yang mutakhir dalam mengatasi tantangan global dengan mengadopsi *remote teaching and learning* [8]. Penelitian menunjukkan bahwa *MOOCs* signifikan dalam mengatasi ketidakseimbangan gender perempuan di bidang *Science, Technology, Engineering and Mathematics (STEM)*, dengan tingkat keberhasilan 38% dibandingkan dengan kuliah konvensional yang hanya 20% [9]. Di Malaysia, *MOOCs* juga diterapkan dalam pelatihan *English for Occupational Purposes (EOP)*, memungkinkan belajar mandiri sesuai kecepatan masing-masing peserta [10]. Penelitian lain menunjukkan bahwa *MOOCs* dapat meningkatkan keterampilan digital karyawan sebesar 45% [11].

Meskipun *MOOCs* menawarkan banyak manfaat, pengguna sering kesulitan menavigasi pilihan kursus yang tersedia, menyebabkan tingginya angka *dropout* [12]. Sebagai contoh, *Canvas Network Platform*, sebuah *MOOCs* besutan *Spanish University* memiliki total 866 peserta didik yang mengambil kursus. Namun, tidak lebih dari 20% peserta yang menyelesaikan kursus dan mendapatkan sertifikat [13]. Hal ini disebabkan latar belakang peserta didik yang heterogen membuat kesulitan mengintegrasikan materi yang sesuai [14]. Selain itu, perspektif *MOOCs* yang *lebih learned-centered* terhadap apa yang mau dipelajari juga meningkatkan retensi peserta untuk menggunakan *MOOCs* [15], [16], [17], [18].

Sistem rekomendasi yang efektif diperlukan untuk membantu pengguna menemukan kursus yang paling relevan dan sesuai dengan minat serta kebutuhan pengguna. Penelitian menunjukkan bahwa mengoptimalkan desain *MOOCs* melalui *PMOOCs* (*Personalized MOOCs*) dapat menjadi solusi untuk mengatasi desain *MOOCs* yang kurang humanis [19]. *PMOOCs* dapat diwujudkan dalam bentuk *E-Learning Recommender System (ERS)* untuk menemukan kombinasi kursus terbaik berdasarkan pengguna [20]. Adapun metode yang umum

digunakan adalah *Collaborative Filtering (CF)* dan *Content Based Filtering (CBF)* [21]. Namun, banyak penelitian yang menggunakan metode ini menghadapi permasalahan seperti: (1) *Sparsity*, yaitu kondisi di mana *rating* yang diperoleh dari pengguna tidak merata dan hanya sedikit pasangan pengguna yang memberikan *rating* pada *item* yang sama. Kelangkaan data ini menyulitkan algoritma similaritas untuk berfungsi dengan baik [22]; dan (2) *Cold Start*, yaitu kondisi di mana pengguna baru masuk ke dalam sistem, menyebabkan hasil rekomendasi menjadi bias karena kurangnya informasi mengenai preferensi pengguna tersebut [23].

Maka dari itu, penelitian ini mengembangkan sistem rekomendasi menggunakan teknik *Weighted Hybrid*, yang menggabungkan *Non-Negative Matrix Factorization (NMF)* dari *Collaborative Filtering* dan *Keyword-Based Filtering (KBF)* dari *Content-Based Filtering*. Teknik *Weighted Hybrid* dapat mengatasi masalah *cold start* dan skalabilitas dengan menggabungkan kelebihan masing-masing metode untuk menutupi kekurangannya [24]. Teknik *Non-Negative Matrix Factorization (NMF)* efektif mengatasi permasalahan *sparsity data*, dengan batasan bahwa semua nilai dalam matriks harus tidak kurang dari nol [25]. Teknik ini cocok untuk penelitian ini karena salah satu variabel yang digunakan adalah *rating* pengguna yang memiliki nilai minimum nol [25]. *NMF* juga memiliki kompleksitas komputasi yang rendah dan relatif mudah diimplementasikan [26]. Sementara itu, *Keyword-Based Filtering (KBF)* memberikan rekomendasi berdasarkan preferensi pengguna dan efektif digunakan bersama dengan *Collaborative Filtering*, karena tetap berfungsi meskipun pengguna tidak memberikan *rating* [27].

Tujuan penelitian ini adalah mengimplementasikan metode *Weighted Hybrid* yang memadukan *NMF* dan *KBF* ke dalam sistem rekomendasi *MOOCs* untuk menghasilkan rekomendasi yang optimal. Hasil rekomendasi diuji menggunakan *Mean Average Precision (MAP)*, metrik evaluasi yang umum digunakan dalam penelitian sistem rekomendasi [28]. Terakhir, hasil *MAP* dikomparasikan dengan penelitian serupa untuk mendapatkan evaluasi yang komprehensif.

*Recommender System* dapat didefinisikan sebagai sistem yang dapat memberikan rekomendasi *item* yang paling sesuai kepada pengguna dengan memprediksi minat dan preferensi pengguna terhadap *item* tersebut [29], [30]. Prediksi ini didasarkan pada informasi terkait *item*, pengguna, dan interaksi antara *item* dan pengguna [29]. Tujuan utama dari *Recommender System* adalah untuk menyajikan hanya *item* yang relevan dengan pengguna, agar dapat mengurangi paparan informasi yang tidak relevan kepada pengguna [29].

*Collaborative Filtering (CF)* adalah salah satu metode yang banyak digunakan di dalam

*Recommender System* [30]. Metode ini menghasilkan rekomendasi dengan mempertimbangkan interaksi orang lain di dalam suatu sistem [29], [30]. Metode ini secara lumrah dibagi menjadi dua pendekatan [30], antara lain: (1) *Memory-Based CF*, pendekatan ini bergantung pada kesamaan antar pengguna. Sederhananya, pengguna yang memiliki banyak kesamaan berarti memiliki minat yang sama juga [30]; dan (2) *Model-Based CF*, pendekatan ini bergantung pada kalkulasi kesamaan antar *item* yang berbeda. Sehingga, rekomendasi dihasilkan dari kalkulasi model prediksi [30].

Berbeda dengan metode *CF* yang mengandalkan interaksi antar pengguna terhadap *item*, *Content-Based Filtering (CBF)* adalah metode yang mengandalkan analisis konten untuk menciptakan rekomendasi [29], [30]. Konten yang dianalisis dapat berupa deskripsi *item* dan profil pengguna [29]. Prinsip dasar dari *CBF* adalah kalkulasi kesamaan antara konten terkait dengan konten yang dibandingkan [31].

Beberapa penelitian telah menerapkan *Collaborative Filtering (CF)* untuk memberikan rekomendasi yang lebih baik kepada pengguna. Misalnya, Erlangga & Sutrisno [32] menggunakan *CF* untuk membantu konsumen memilih salon kecantikan yang sesuai berdasarkan ulasan pelanggan lain. Sementara itu, Februariyanti dkk. [33] menerapkan *CF* untuk merekomendasikan produk mebel pada toko online, sehingga pelanggan tidak perlu datang langsung ke toko. Kedua penelitian ini menunjukkan bahwa *CF* dapat efektif dalam memberikan rekomendasi yang relevan.

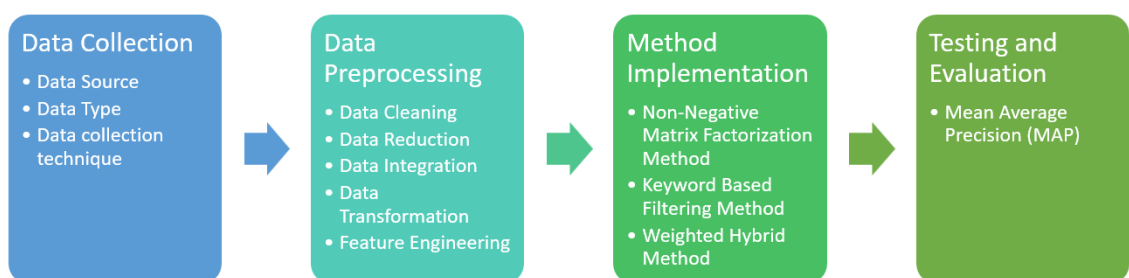
*Hybrid method* adalah metode yang menggabungkan dua atau lebih metode untuk mengatasi keterbatasan dan kekurangan dari masing-masing metode agar meningkatkan kualitas rekomendasi yang diberikan [31]. Terdapat tujuh macam teknik di dalam *hybrid method*, antara lain [28], [34]: (1) *Weighted*, yaitu teknik pemberian bobot terhadap masing-masing metode; (2) *Switching*, yaitu teknik pergantian metode yang diterapkan berdasarkan keadaan sekarang; (3) *Mixed*, yaitu teknik yang menampilkan hasil dari berbagai rekomendasi secara bersamaan; (4) *Feature Combination*, yaitu penggunaan informasi dari suatu metode untuk digunakan kembali pada metode yang

lain; (5) *Feature Augmentation*, yaitu penggunaan fitur dari metode yang berbeda menjadi satu algoritma tunggal; (6) *Cascade*, yaitu teknik mengambil hasil rekomendasi metode yang lain untuk diolah ke dalam metode berikutnya; dan (7) *Meta-Level*, yaitu penggunaan keseluruhan model dari metode pertama untuk digunakan sebagai input bagi metode berikutnya.

Beberapa penelitian telah menggabungkan metode *content-based filtering* dan *collaborative filtering (metode hibrid)* untuk meningkatkan kualitas rekomendasi. Misalnya, Wijaya & Alfian [34] menggunakan metode hibrid untuk membantu pengguna memilih laptop yang sesuai, sementara Subroto dkk. [35] menerapkan metode hibrid untuk merekomendasikan materi pembelajaran kepada siswa. Penelitian Widjaja & Palit [28] juga menunjukkan bahwa metode hibrid memberikan hasil yang lebih baik dalam merekomendasikan buku di perpustakaan. Hasil penelitian-penelitian ini menunjukkan bahwa metode hibrid dapat mengatasi kekurangan dari masing-masing metode individu dan memberikan rekomendasi yang lebih akurat.

Berdasarkan kelima penelitian terdahulu, dapat disimpulkan bahwa penelitian pertama dan kedua menggunakan metode yang sama dengan teknik yang berbeda [32], [33]. Namun, baik antara teknik *user-based* ataupun *item-based* memiliki kekurangan yang sama dikarenakan kesamaan pendekatan yang digunakan. *Collaborative Filtering* terkhusus pada pendekatan *memory-based* memiliki permasalahan mendasar yaitu *sparsity* dan *cold start* [22], [23]. Menanggapi permasalahan *cold start* ini, solusi yang dapat diterapkan adalah metode *hybrid* yang digunakan dalam penelitian ketiga hingga kelima [28], [34], [35]. Metode ini dalam beberapa penelitian terbukti mampu mengatasi atau setidaknya mereduksi ketidakakuratan rekomendasi yang diakibatkan oleh *cold start problem* [36]. Sedangkan untuk permasalahan *sparsity*, kelima penelitian tersebut tidak menggunakan pendekatan *model-based* yang mampu memberikan rekomendasi yang lebih akurat walaupun data yang diolah sangat sedikit [24].

## 2. METODE PENELITIAN



Gambar 1. Desain Penelitian

Dalam melaksanakan penelitian, diperlukan tahapan yang sistematis agar mampu mencapai tujuan

penelitian dengan baik. Desain penelitian ini mengacu pada beberapa penelitian terdahulu yang

meneliti topik serupa [37], [38], [39], [40], [41]. Penelitian yang membutuhkan data memerlukan tahapan pengumpulan data dan pra-pemrosesan data. Selanjutnya, metode yang menjadi fokus utama penelitian diterapkan. Terakhir, tahap pengujian dan evaluasi dilakukan untuk mengukur keberhasilan dari metode *recommender system* yang digunakan.

Pada Gambar 1, langkah awal dari penelitian ini adalah mengumpulkan data. Data yang dikumpulkan salah satunya adalah *rating* pengguna terhadap kursus di dalam *MOOCs*. Selanjutnya, data ini memasuki tahap pra-pemrosesan yang bertujuan untuk mempersiapkan data yang akan diolah ke tahap berikutnya. Lalu, metode diimplementasikan terhadap data-data tersebut. Akhirnya, hasil dari metode ini kemudian diuji dan dievaluasi untuk diskusi lebih lanjut.

### 2.1. Data Collection

*Dataset* yang digunakan dalam penelitian ini diperoleh dari Kaggle (<https://www.kaggle.com/datasets/koutetsu/udemy-courses-and-reviews/data>). Data ini berisi informasi dan penilaian kursus dari *platform* *MOOCs* "Udemy" selama periode tahun 2021. Terdapat dua *dataset* yang digunakan. *Dataset* pertama berjudul *courses.yaml* dan *dataset* kedua berjudul *reviews.yaml*. Adapun informasi *dataset* ini dijabarkan pada Tabel 1 dan Tabel 2.

Tabel 1. Informasi *Dataset* Pertama

No.	Informasi	Deskripsi
1	<i>Range Index</i>	98000 <i>entries</i> , 0 to 97999
2	<i>Columns</i>	103 <i>entries</i> , <i>course_id</i> to <i>objectives/31</i>
3	<i>Data Types</i>	bool(3), float64(3), int64(15), object(82)
4	<i>Memory Usage</i>	7.5+ MB

Pada Tabel 1, *dataset* pertama memiliki 98,000 entri dengan indeks mulai dari 0 hingga 97,999. Adapun kolom terdiri dari 103 entri, mulai dari kolom *course\_id* hingga *objectives/31*. Tipe data pada *dataset* ini terdiri dari boolean sebanyak tiga kolom, float64 sebanyak tiga kolom, int64 sebanyak 15

kolom dan object sebanyak 82 kolom. Ukuran *dataset* ini berkisar 7.5 *Megabyte*.

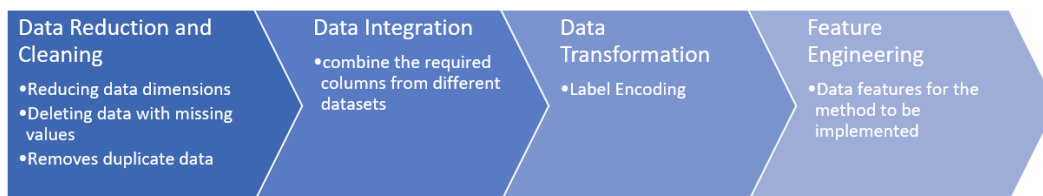
Tabel 2. Informasi *Dataset* Kedua

No.	Informasi	Deskripsi
1	<i>Range Index</i>	98000 <i>entries</i> , 0 to 97999
2	<i>Columns</i>	103 <i>entries</i> , <i>course_id</i> to <i>objectives/31</i>
3	<i>Data Types</i>	bool(3), float64(3), int64(15), object(82)
4	<i>Memory Usage</i>	7.5+ MB

Pada Tabel 2, *dataset* kedua memiliki 58.408 entri dengan indeks mulai dari 0 hingga 58,407. Adapun kolom terdiri dari tujuh entri, mulai dari kolom *course\_id* hingga *review\_modified*. Tipe data pada *dataset* ini terdiri dari float64 sebanyak dua kolom, int64 sebanyak dua kolom dan object sebanyak tiga kolom. Ukuran *dataset* ini berkisar 3,5 *Megabyte*.

### 2.2. Data Preprocessing

Setelah tahap pengumpulan data, *dataset* kemudian masuk ke dalam tahap pra-pemrosesan. Pra-pemrosesan data bertujuan agar data yang digunakan berkualitas baik dan sesuai dengan kebutuhan penelitian [42], [43]. Sesuai Gambar 2, adapun tahapan-tahapan pra-pemrosesan data dalam penelitian ini meliputi *data reduction and cleaning*, *data integration*, *data transformation* dan *feature engineering*. Tahap *data reduction* dilakukan dengan mereduksi kolom-kolom yang tidak relevan [44]. Tahap *data cleaning* dilakukan dengan cara menghapus *missing value* dan duplikasi di dalam *dataset* [43], [45], [46]. Setelah itu, tahap *data integration* dilakukan dengan cara menggabungkan kolom-kolom yang diperlukan dari dua *dataset* ke dalam bentuk satu *dataframe* untuk mengefisienkan pengolahan data [43], [47]. Selanjutnya, tahap *data transformation* dilakukan melalui proses *label encoding* dengan cara mengubah bentuk data kategorikal menjadi bentuk numerik untuk menandai data dari kategori yang sama [44], [48]. Terakhir, tahap *feature engineering* dilakukan dengan menyiapkan fitur-fitur yang dapat mengefisienkan kinerja di tahap berikutnya [49].



Gambar 2. Tahapan *Data Preprocessing*

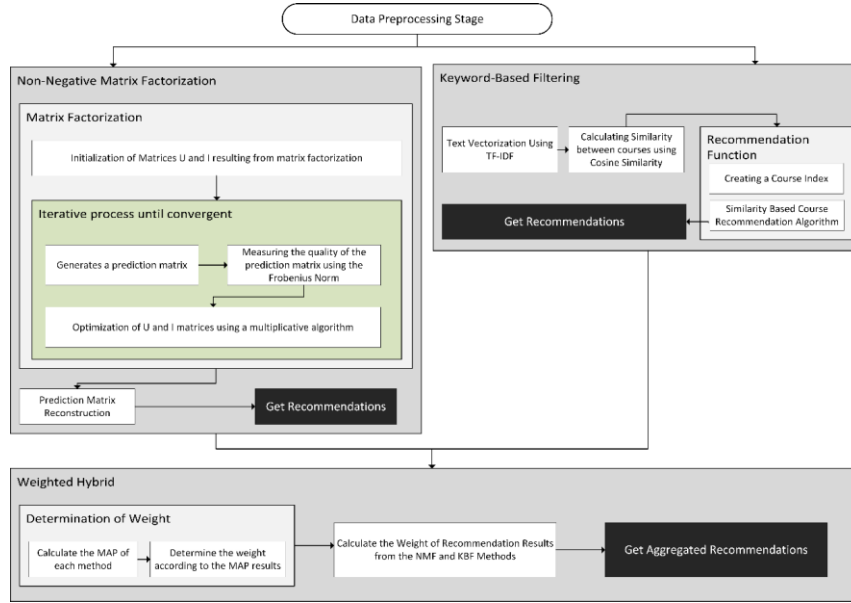
### 2.3. Method Implementation

Dalam penelitian ini, metode yang diimplementasikan adalah metode *Non-Negative Matrix Factorization (NMF)*, *Keyword-Based Filtering* dan *Weighted Hybrid*. Tahap implementasi

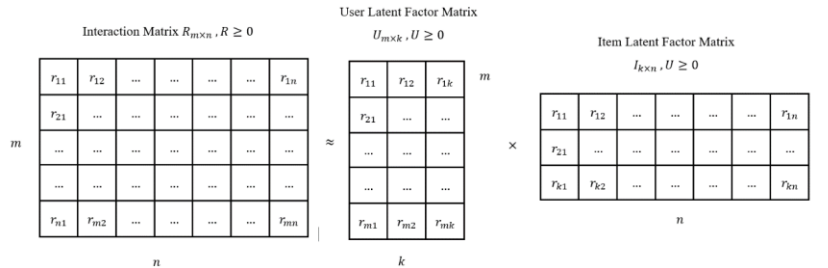
metode dalam penelitian ini digambarkan dalam Gambar 3.

*Non-Negative Matrix Factorization (NMF)* adalah salah satu teknik di dalam metode *Collaborative Filtering (CF)* yang digunakan untuk memprediksi matriks *rating* [26]. Sebagai bagian dari pendekatan *model-based*, faktorisasi matriks

memungkinkan memberikan keakuratan rekomendasi yang baik [26].



Gambar 3. Tahapan Method Implementation



Gambar 4. Proses Faktorisasi Matriks

Gambar 4 menjelaskan bahwa matriks interaksi merupakan matriks yang rumpang dengan banyak nilai yang kosong. Sebuah matriks interaksi  $R_{m \times n}, R \geq 0$  memiliki elemen  $r_{mn}$  yang mendeskripsikan preferensi pengguna  $m$  terhadap item  $n$  dan peringkat matriks  $k$ . Selanjutnya, matriks ini difaktorisasi menjadi dua non-negatif matriks, yaitu matriks faktor laten pengguna  $U_{m \times k}$  dan matriks faktor laten item  $I_{k \times n}$  [50], [51]. Matriks  $U$  dan  $I$  ini berperingkat lebih rendah daripada matriks  $R$  dikarenakan  $k \leq \min(n, m)$  [26].

$$R \approx \hat{R} \approx UI \quad (1)$$

Pada persamaan (1), Matriks  $U$  dan  $I$  mendeskripsikan faktor laten dari pengguna dan item yang melalui operasi *matrix product* mampu menghampiri matriks asal  $R$  dalam bentuk matriks prediksi  $\hat{R}$  [26]. Dikarenakan matriks  $U$  dan  $I$  dinisialisasi dengan elemen matriks non-negatif yang acak, Frobenius Norm digunakan untuk mengukur seberapa baik matriks  $\hat{R}$  dalam merekonstruksi matriks  $R$  [52].

$$\|R - \hat{R}\|_F = \sqrt{\sum_{m,n} (R_{mn} - \hat{R}_{mn})^2} \quad (2)$$

Pada persamaan (2), matriks  $\hat{R}$  dianggap paling mendekati matriks  $R$  apabila  $\|R - \hat{R}\|_F$  berada pada kondisi konvergen. Kondisi konvergen yang akan digunakan pada penelitian ini adalah ketika perubahan nilai Frobenius Norm antara dua iterasi dibawah ambang batas  $1 \times 10^{-4}$ . Untuk mencapai kondisi konvergen, diperlukan iterasi untuk merubah elemen-elemen pada matriks  $U$  dan  $I$  melalui algoritma multiplikatif berikut [53]:

$$U_{mk} \leftarrow U_{mk} \frac{(R I^T)_{mk}}{(U I I^T)_{mk}} \quad (3)$$

$$I_{kn} \leftarrow I_{kn} \frac{(U^T R)_{kn}}{(U^T U I)_{kn}} \quad (4)$$

Setelah kondisi konvergen telah tercapai, selanjutnya matriks faktor laten  $U$  dan  $I$  direkonstruksi menjadi matriks *rating* agar dapat digunakan untuk mendapatkan rekomendasi kursus bagi pengguna (lihat persamaan (1)). Pengguna  $m$  akan mendapatkan *Top-N Recommendations* berdasarkan *rating* yang telah diprediksi dan diurutkan dari *rating* yang tertinggi.

Tidak seperti metode *NMF*, metode *Keyword-Based Filtering* diimplementasikan untuk

mendapatkan rekomendasi kursus berdasarkan preferensi pengguna [54]. Terkhusus pada penelitian ini, metode ini digunakan untuk menghasilkan rekomendasi berdasarkan kursus yang diikuti pengguna dengan melihat similaritas antar kursus. Adapun langkah awal dari metode ini adalah menghasilkan matriks profil kursus menggunakan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*) sebagai berikut [54]:

$$TF(t, d) = \frac{f_{t,d}}{N_d} \quad (5)$$

$$IDF(t, D) = \log\left(\frac{N}{n_t}\right) \quad (6)$$

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (7)$$

Pada persamaan (5), fungsi *TF* mengukur seberapa sering sebuah kata *t* muncul dalam sebuah judul kursus *d* (di mana  $f_{t,d}$  adalah jumlah kemunculan kata *t* dalam sebuah judul kursus *d* dan  $N_d$  adalah total jumlah kata dalam judul kursus *d*). Fungsi *IDF* mengukur frekuensi kursus yang mengandung kata *t* di dalam korpus judul kursus *D* (dimana  $N$  adalah total judul kursus di dalam korpus dan  $n_t$  adalah jumlah kursus *d* yang mengandung kata *t*) (lihat persamaan (6)). Pada persamaan (7), nilai matriks profil kursus didapatkan dari mengalikan antara nilai *TF* dan *IDF*. Selanjutnya, nilai *TF-IDF* digunakan untuk menghitung *Cosine Similarity* antara kursus yang diikuti pengguna dan seluruh kursus yang ada [27]. *Cosine Similarity* adalah ukuran kesamaan antara dua vektor non-nol [27]. Adapun *Cosine Similarity* didapatkan melalui persamaan berikut:

$$\cos \alpha = \frac{A \cdot B}{\|A\| \|B\|} \quad (8)$$

Pada persamaan (8),  $\cos \alpha$  adalah nilai *cosinus* antara dua vektor *TF-IDF* dari kursus *A* dan *B*. Selanjutnya, nilai magnitudo vektor *A* dan *B* ( $\|A\|$ ,  $\|B\|$ ) didapatkan melalui persamaan berikut:

$$\|A\| = \sqrt{\sum_{i=1}^n A_i^2} \quad (9)$$

$$\|B\| = \sqrt{\sum_{i=1}^n B_i^2} \quad (10)$$

Magnitudo vektor *A* dan *B* adalah hasil akar kuadrat dari jumlah kuadrat dari setiap elemen vektor. Setelah nilai *Cosine Similarity* didapatkan, rekomendasi dihasilkan dalam bentuk *Top-N Recommendations* berdasarkan urutan nilai *Cosine Similarity* tertinggi di antara seluruh kursus terhadap indeks kursus yang diikuti pengguna.

Selanjutnya, metode *Weighted Hybrid* diimplementasikan berdasarkan hasil rekomendasi metode sebelumnya. *Weighted Hybrid* menghasilkan

rekomendasi dari gabungan hasil rekomendasi metode *NMF* dan *KBF* pada pembobotan tertentu [28]. Pada penelitian ini, bobot masing-masing metode diberikan berdasarkan nilai *Mean Average Precision* (*MAP*) [54]. Sehingga, misalkan  $MAP_{NMF}$  dan  $MAP_{KBF}$  merupakan nilai *MAP* dari hasil rekomendasi metode *NMF* dan *KBF* maka bobot masing-masing metode dapat diraih dengan persamaan berikut [55], [56]:

$$\omega_{NMF} = \frac{MAP_{NMF}}{MAP_{NMF} + MAP_{KBF}} \quad (11)$$

$$\omega_{KBF} = \frac{MAP_{KBF}}{MAP_{NMF} + MAP_{KBF}} \quad (12)$$

Berdasarkan persamaan (11) dan (12), matriks prediksi  $\hat{R}_\omega$  dari metode *Weighted Hybrid* dapat diraih dengan persamaan berikut [55]:

$$\hat{R}_\omega = \omega_{NMF} \cdot \hat{R}_{NMF} + \omega_{KBF} \cdot \hat{R}_{KBF} \quad (13)$$

Dari persamaan (13), hasil rekomendasi dapat direkonstruksi ulang dengan mengurutkan item kursus yang telah diberi pembobotan dari metode *NMF* dan *KBF* dalam bentuk *Top-N Recommendation* dengan urutan dari yang tertinggi.

## 2.4. Testing and Evaluation

Pada penelitian ini, *Mean Average Precision* (*MAP*) digunakan untuk mengukur kinerja metode dalam menghasilkan rekomendasi [54]. *MAP* cocok digunakan untuk metode yang menghasilkan urutan peringkat dalam bentuk *list*, di mana *item* rekomendasi pada penelitian ini dapat *hit* (*item* relevan) atau *miss* (*item* non-relevan) [54]. Adapun langkah awal dari *MAP* adalah menghitung *precision at k* or  $P(k)$  untuk pengguna *u* dengan persamaan berikut:

$$P(k) = \frac{\sum_{i=1}^k Rel(i)}{k} \quad (14)$$

Pada persamaan (14),  $P(k)$  diraih dengan menghitung jumlah *item* relevan hingga posisi *k* ( $Rel(i)$  adalah fungsi indikator yang bernilai 1 jika *item* relevan dan 0 jika tidak) dibagi dengan posisi *k*. Untuk mengetahui *item* rekomendasi bernilai relevan atau tidak, diperlukan daftar *item* relevan sebagai acuan. Pada penelitian ini, daftar *item* relevan diraih dari daftar kursus dengan tingkat kesulitan yang sama dengan kursus yang telah di-*rating* oleh pengguna. Selanjutnya, nilai  $P(k)$  pada *item* relevan digabungkan untuk mendapatkan *Average Precision* (*AP*) untuk setiap pengguna *u* [54]. Misalkan,  $|R|$  merupakan jumlah *item* relevan pada hasil rekomendasi, maka *Average Precision* dapat diraih melalui persamaan berikut [54]:

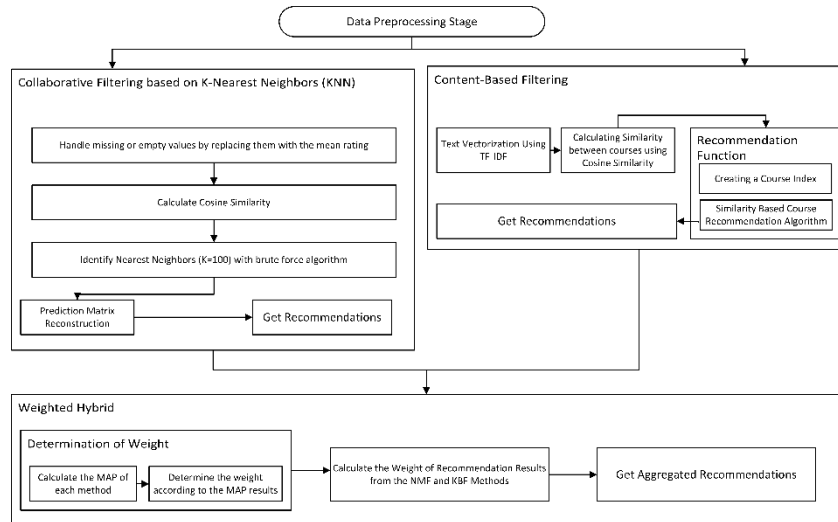
$$AP = \frac{1}{|R|} \sum_{k=1}^n P(k) \cdot rel(k) \quad (15)$$

Selanjutnya, misalkan  $|U|$  adalah jumlah pengguna, maka nilai  $AP$  setiap pengguna digabungkan untuk mendapatkan nilai rerata dari *Average Precision (MAP)* dengan persamaan berikut [54]:

$$MAP = \frac{1}{|U|} \sum_{i=1}^U AP_u \tag{16}$$

Hasil  $MAP$  dari *Weighted Hybrid* kemudian dievaluasi kinerjanya dengan melakukan komparasi

dengan penelitian yang serupa [28]. Penelitian tersebut menggunakan metode *Weighted Hybrid* dengan mengkombinasikan *Collaborative Filtering* berbasis *K-Nearest Neighbors (kNN)* dan *Content Based Filtering*. Untuk evaluasi penelitian ini, metode yang digunakan pada penelitian [28] diimplementasikan ulang terhadap dataset penelitian ini. Dengan beberapa penyesuaian, berikut alur metode yang digunakan sebagai pembandingan penelitian ini:



Gambar 5. Alur Metode Pembandingan Penelitian

Secara garis besar, metode pembandingan yang digunakan memiliki alur yang mirip dengan penelitian ini. Hanya saja, metode penyusunnya berbeda. Metode *CBF* yang digunakan berbasis *KNN* dengan jumlah *neighbors* yang digunakan adalah 100 [28]. Sedangkan metode *CBF* yang digunakan adalah metode yang sama dengan *Keyword Based Filtering*, yaitu menghitung kesamaan pada judul kursus.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Dataset

Adapun sampel dari *dataset* penelitian ini dijabarkan pada Tabel 3 dan Tabel 4.

Tabel 3. Isi *Dataset* Pertama

course_id	url	title	...	objectives/29	objectives/30	objectives/31
3701926	/course/letting-go-during-a-life-crisis/	How to Burn through Emotions	...	NaN	NaN	NaN
3701752	/course/scrum-product-owner-1-prepare-for-the-...	Scrum Product Owner 1. prepare for the PSPO I...	...	NaN	NaN	NaN
...	...	...	...	...	...	...
3684424	/course/sap-fiori-learn-mm-in-sap-fiori-from-s...	SAP Fiori- Learn MM in SAP Fiori from Scratch...	...	NaN	NaN	NaN
3684408	/course/mobile-device-farm-st...	Mobile Testing Device Lab...	...	NaN	NaN	NaN

Tabel 4. Isi *Dataset* Kedua

course_id	review_id	review_title	...	review_rating	review_created	review_modified
3701752	95068444	NaN	...	5.0	2021-09-09T14:56:26Z	2021-09-10T16:03:43Z
3701752	78858128	NaN	...	5.0	2021-02-27T09:25:46Z	2021-02-28T19:08:43Z
...	...	...	...	...	...	...
3684408	76052120	NaN	...	5.0	2021-01-23T23:36:11Z	2021-01-25T15:05:00Z
3684408	75740872	NaN	...	5.0	2021-01-20T02:54:33Z	2021-01-21T15:04:59Z

#### 3.2. Data Preprocessing

Pada tahap *data preprocessing*, *dataset* direduksi dimensinya dengan mengurangi kolom-kolom yang tidak digunakan dalam penelitian.

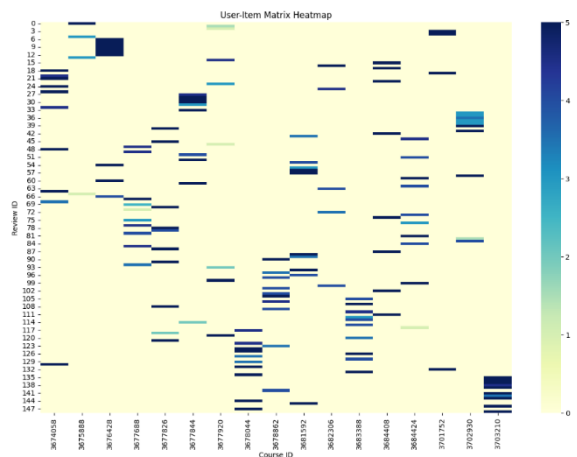
Setelah direduksi, *dataset* dibersihkan dari baris yang memiliki nilai kosong dan duplikat. Sehingga, *dataset* menjadi lebih efektif. Adapun kolom yang dipertahankan pada *dataset* pertama yaitu: (1) *course\_id*, kolom ini berisi tanda pengenal kursus; (2) *published\_title*, kolom ini berisi judul kursus; dan (3)

instructional\_level\_simple, kolom ini berisi tingkat kesulitan. Sedangkan pada *dataset* kedua yaitu: (1) *course\_id*; (2) *review\_id*, kolom ini berisi tanda pengenalan pengguna; dan (3) *review\_rating*, kolom ini berisi penilaian pengguna.

Pada tahap *data integration*, *dataframe* baru dibentuk dengan menggabungkan kolom-kolom dari dua *dataset* agar menjadi terpadu dan mudah diolah. Adapun dua *dataset* ini digabung ke berdasarkan satu kolom acuan yang sama, yaitu *course\_id*. Sehingga, *dataframe* ini (*merged\_df*) memiliki kolom sebagai berikut: (1) *course\_id*, (2) *review\_id*, (3) *published\_date* dan (4) *instructional\_level\_simple*. Kolom *review\_id* merupakan kolom yang memiliki data kategorikal, artinya *user* bisa saja memberikan *rating* ke lebih dari satu kursus.

Untuk memudahkan pengelompokan data, dilakukan *label encoding* untuk mengganti *review\_id* dengan label *integer*. Dalam penelitian ini, *review\_id* diberi label mulai dari 0. Sehingga, diketahui terdapat 148 *user* yang memberikan *rating*. Selanjutnya, tahap *feature engineering* digunakan untuk menyiapkan fitur yang mempermudah implementasi metode. Adapun untuk metode *NMF*, Matriks *user-item* dibuat untuk melihat interaksi antara pengguna dan kursus dalam konteks pemberian *rating*. Matriks ini memiliki baris berdasarkan jumlah *user* dan kolom berdasarkan jumlah kursus.

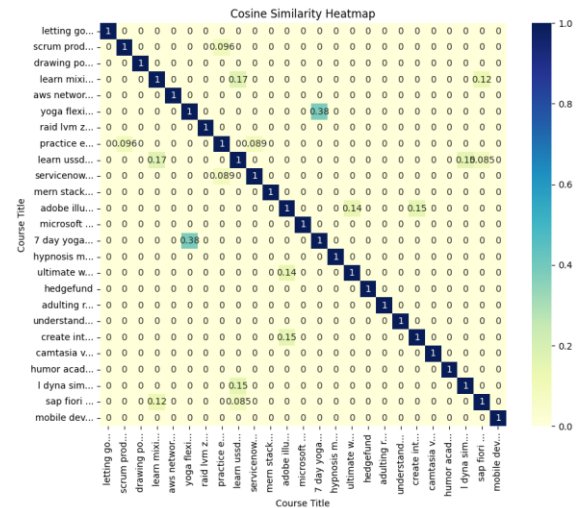
Gambar 6 merupakan *heatmap* interaksi antara pengguna (*review\_id*) dan item (*course\_id*) dalam sebuah matriks. Visualisasi ini menunjukkan kepopuleran dari suatu kursus dilihat dari interaksi *rating*-nya. Sedangkan warna sel menunjukkan tingkat *rating* kursus. Dari *heatmap* ini diketahui bahwa sebagian besar pengguna ternyata hanya memberikan *rating* kepada satu kursus. Sehingga, kondisi data dalam penelitian ini tergolong jarang/langka, yang artinya terdapat sangat sedikit informasi yang dapat diolah.



Gambar 6. User-Item Matrix Heatmap

Adapun untuk metode *KBF*, *Natural Language Processing (NLP)* diterapkan untuk kolom *published\_title* dikarenakan kolom ini berisi *natural language* yang kemudian dapat ditelaah similaritas

antar kursus dari judulnya. Adapun *NLP* yang digunakan dalam penelitian ini sebagai berikut: (1) *Text preprocessing*, yaitu menghapus karakter khusus dan mengubah karakter menjadi *lowercase*; (2) *Tokenisasi*, yaitu proses memecah teks menjadi kata-kata individu; (3) *Stop words removal*, yaitu proses menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis; dan (4) *Lematisasi*, yaitu proses mengubah kata-kata menjadi bentuk dasarnya.



Gambar 7. Cosine Similarity Heatmap

Setelah tahap *NLP*, proses dilanjutkan untuk menghasilkan matriks *TF-IDF*. Matriks ini menggambarkan pentingnya kata dalam suatu judul kursus dalam kumpulan judul kursus. Selanjutnya, skor *TF-IDF* ini digunakan untuk mengukur kesamaan kosinus antar kursus. Gambar 7 mengilustrasikan matriks kemiripan kosinus antar judul kursus. Setiap sel dalam *heatmap* mewakili tingkat kemiripan antara dua judul kursus, di mana warna dan angka di dalamnya menunjukkan seberapa mirip kedua judul tersebut berdasarkan skor *TF-IDF*. Dari *heatmap* ini dapat dilihat bahwa sangat sedikit kursus yang memiliki kemiripan dengan kursus yang lain. Hal ini dapat mengindikasikan hasil rekomendasi yang mungkin kurang akurat.

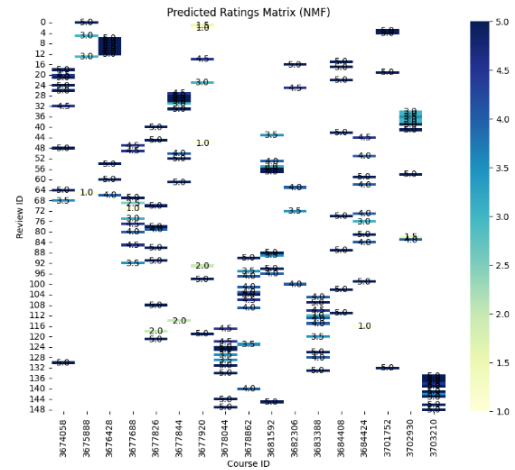
### 3.3. Hasil

Pada metode *NMF*, pembentukan matriks prediksi dilakukan dengan membuat model *NMF* menggunakan matriks interaksi *user-item*. Model *NMF* ini menggunakan faktor laten ( $n = 20$ ).

Gambar 8 menjabarkan matriks prediksi hasil dari *training model NMF*. Pada beberapa sel terjadi pembentukan prediksi *rating* yang beragam, mulai dari kemunculan *rating* sangat kecil hingga *rating* yang hampir sama dengan *rating* pada matriks interaksi. Iterasi terakhir ini menunjukkan kondisi konvergensi yang telah tercapai. Sehingga matriks prediksi ini memiliki nilai Frobenius Norm yang hampir sama dengan matriks interaksi. Algoritma rekomendasi selanjutnya menggunakan matriks



prediksi ini untuk mendapatkan *Top-5 recommendations* bagi pengguna dengan mengurutkan *rating* dari yang tertinggi. Adapun sampel hasil rekomendasi dari metode ini sebagai berikut.



Gambar 8. *Predicted Ratings Heatmap*

Tabel 5. Hasil Rekomendasi dari Metode *NMF*

user_id	Order of Recommendations				
	1	2	3	4	5
0	raid-lvm-zfs-iscsi-protocol-fog-pxe-on-centos-linux-8	yoga-for-flexibility-from-zero-to-hero-in-15-minutes-a-day	hypnosis-meditation-wellbeing-ithc-mastery-certificate	adobe-illustrator-perfect-beginner-course	practice-examsprofessional-scrum-master-ii-psm-2-psm-ii-n
56	camtasia-video-editing-youtube-animation-	yoga-for-flexibility-from-zero-to-he	raid-lvm-zfs-iscsi-protocol-fog-	adobe-illustrator-perfect-beg	hypnosis-meditation-wellbeing-
148	aws-networking-amazon-vmc-aws-...	yoga-for-flexibility-from-zero-to-hero-...	raid-lvm-zfs-iscsi-protocol-fog-pxe-on-...	adobe-illustrator-perfect-beginner-...	hypnosis-meditation-wellbeing-ithc...

Metode *KBF* diimplementasikan dengan menggunakan matriks *cosine similarity* (lihat Gambar 5). Algoritma rekomendasi menghasilkan rekomendasi berdasarkan judul kursus yang telah *rating* pengguna. Maka, tingkat kemiripan terhadap

judul kursus ini lalu diurutkan dari yang tertinggi untuk mendapatkan *Top-5 recommendations*. Adapun sampel rekomendasi dari metode ini sebagai berikut.

Tabel 6. Hasil Rekomendasi dari Metode *KBF*

user_id	Order of Recommendations				
	1	2	3	4	5
0	letting-go-during-a-life-crisis	scrum-product-owner-1-prepare-for-the-pspo-i-exam	drawing-portraits-the-eye	learn-mixing-fundamental	aws-networking-amazon-vmc-aws-vmc-hybrid-cloud
56	letting-go-during-a-life-crisis	scrum-product-owner-1-prepare-for-the-pspo-i-exam	drawing-portraits-the-eye	learn-mixing-fundamental	aws-networking-amazon-vmc-aws-vmc-hybrid-cloud
148	letting-go-during-a-life-crisis	scrum-product-owner-1-prepare-for-the-pspo-i-exam	drawing-portraits-the-eye	learn-mixing-fundamental	yoga-for-flexibility-from-zero-to-hero-in-15-m...

Pada metode *Weighted Hybrid*, pembobotan diperlukan untuk menghasilkan rekomendasi yang proporsional. Pembobotan untuk metode *NMF* dan *KBF* didasarkan atas skor *MAP* dari setiap metode (lihat persamaan (13)). Adapun nilai *MAP* dan pembobotan tiap metode sebagai berikut:

<i>Keyword Filtering</i>	<i>Based</i>	0.0905	0.25
--------------------------	--------------	--------	------

Tabel 7. *Method Weight*

Method	MAP Score	Method Weight	Total Weight
<i>Non-Negative Matrix Factorization</i>	0.2636	0.75	1

Selanjutnya, *Top-5 recommendation* dari metode *NMF* dan *KBF* digabungkan ke dalam satu list dan diberi pembobotan berdasarkan asal rekomendasi tersebut (persamaan (13)). Hasil rekomendasi gabungan ini kemudian diurutkan ulang. Adapun sampel rekomendasi dari metode ini sebagai berikut.

Tabel 8. Hasil Rekomendasi dari Metode *Weighted Hybrid*

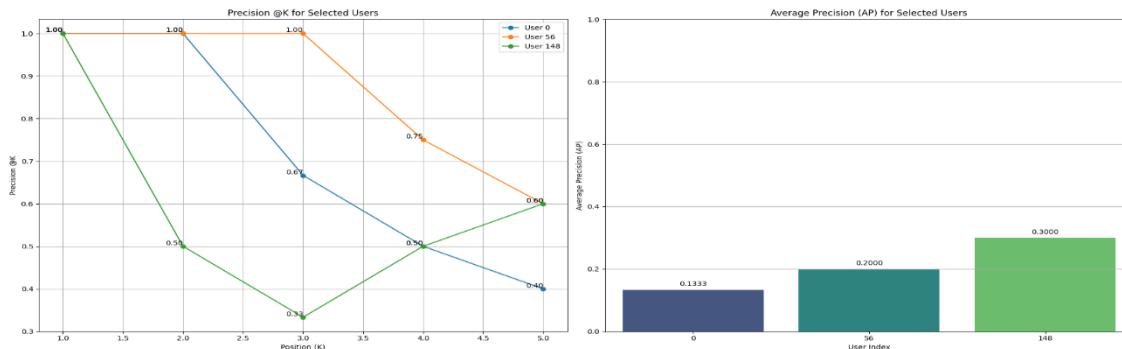
user_id	Order of Recommendations				
	1	2	3	4	5
0	practice-examsprofessional-scrum-master-ii-psm-2-psm-ii-n	yoga-for-flexibility-from-zero-to-hero-in-15-minutes-a-day	hypnosis-meditation-wellbeing-ithc-mastery-certificate	raid-lvm-zfs-iscsi-protocol-fog-pxe-on-centos-linux-8	adobe-illustrator-perfect-beginner-course
56	hypnosis-meditation-wellbeing-ithc-mastery-certificate	yoga-for-flexibility-from-zero-to-hero-in-15-minutes-a-day	raid-lvm-zfs-iscsi-protocol-fog-pxe-on-centos-linux-8	camtasia-video-editing-youtube-animation-urdu	adobe-illustrator-perfect-beginner-course

148	yoga-for-flexibility-from-zero-to-hero-in-15-minutes-a-day	aws-networking-amazon-vpc-aws-vpn-hybrid-cloud	hypnosis-meditation-	raid-lvm-zfs-iscsi-protocol-fog-pxe-on-centos-linux-8	adobe-illustrator-perfect-beginner-course
-----	--	--	----------------------	---	---

#### 4. DISKUSI

Untuk menguji performa dari metode pada penelitian ini, *MAP* digunakan untuk mengukur rerata

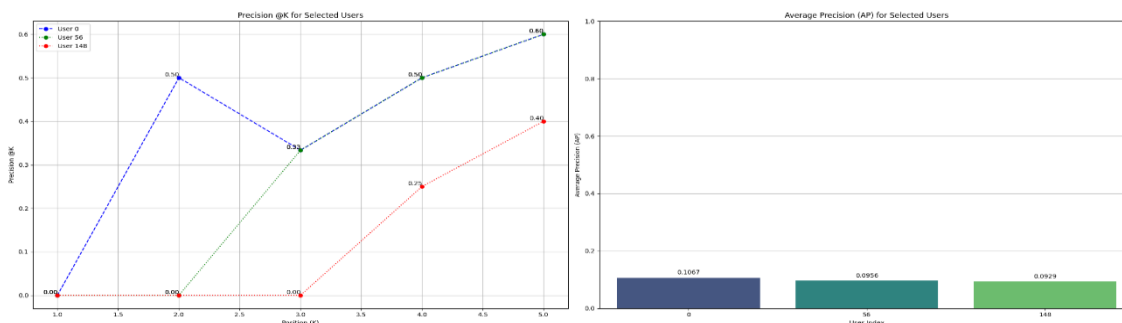
nilai *Average Precision (AP)* seluruh pengguna. Adapun langkah awalnya adalah mendapatkan nilai  $P(k)$  dari pengguna. Kemudian, mendapatkan nilai *AP* pengguna. Terakhir, *MAP* diraih dari rerata nilai *AP* dari seluruh pengguna.



Gambar 9. Precision @k and Average Precision Results

Pada Gambar 9, *Precision @k* pada sampel pengguna menunjukkan nilai *precision* (kemunculan *item* relevan) pada hasil rekomendasi satu teratas hingga lima teratas. Pada posisi  $k = 1$ , tiga sampel menunjukkan nilai *precision* sama dengan satu yang berarti pasti terdapat kursus relevan yang muncul di posisi teratas. Lalu, nilai *precision* menurun seiring dengan jumlah kursus yang direkomendasikan.

Sedangkan, nilai *AP* pada sampel pengguna menunjukkan nilai yang variatif. Pada *user 148*, nilai *AP* sebesar 0.3000 menunjukkan bahwa ketika sistem memberikan rekomendasi, rata-rata 30% dari kursus yang direkomendasikan tersebut relevan bagi pengguna.

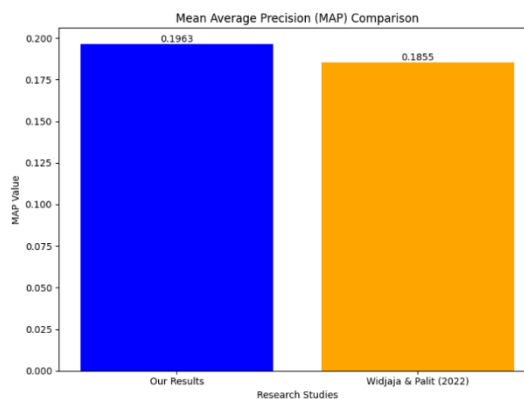


Gambar 10. Precision @k and Average Precision Results from Similar Research

Sedangkan pada penelitian serupa [28], *Precision @k* pada posisi  $k = 1$  menunjukkan nilai *precision* sama dengan nol yang berarti tidak ada kursus relevan yang muncul di posisi teratas. Lalu, nilai *precision* meningkat seiring dengan jumlah kursus yang direkomendasikan. Sedangkan, nilai *AP* pada sampel pengguna menunjukkan nilai yang variatif. Pada *user 148*, nilai *AP* sebesar 0.0929 menunjukkan bahwa ketika sistem memberikan rekomendasi, rata-rata hanya sekitar 9,29% dari kursus yang direkomendasikan tersebut relevan bagi pengguna.

memberikan performa yang lebih baik dalam menghasilkan rekomendasi yang relevan berdasarkan evaluasi menggunakan *MAP*.

Pada Gambar 11, hasil menunjukkan bahwa penelitian ini mencapai nilai *MAP* sebesar 0,1963, sementara penelitian serupa [28] menghasilkan nilai *MAP* sebesar 0,1855. Dengan demikian, dari hasil evaluasi ini, metode pada penelitian ini mampu



Gambar 11. Perbandingan Mean Average Precision (MAP)

## 5. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa metode *Weighted Hybrid* yang menggabungkan metode *NMF* dan *KBF* menunjukkan performa yang unggul. Metode ini berhasil mengatasi masalah *cold start* dan *sparsity*, serta meningkatkan skalabilitas dengan menggabungkan kelebihan dari masing-masing metode.

Untuk penelitian mendatang, beberapa saran ini dapat dipertimbangkan berdasarkan hasil dari penelitian ini. Pertama, penggunaan *dataset* yang lebih lengkap akan sangat meningkatkan evaluasi performa metode *Weighted Hybrid* yang digunakan. Selanjutnya, optimisasi parameter *Weighted Hybrid* juga dapat ditingkatkan dengan meneliti lebih lanjut pengaturan bobot antara *NMF* dan *KBF* untuk mencapai hasil yang lebih optimal. Terakhir, Evaluasi menggunakan metrik tambahan seperti *Mean Absolute Error (MAE)* dan *F1-score* juga perlu dipertimbangkan untuk memberikan pemahaman yang lebih komprehensif terhadap kinerja sistem rekomendasi.

## DAFTAR PUSTAKA

- [1] S. Symonenko, N. Zaitseva, and V. Osadchyi, "Implementation of MOOC platforms into teaching English to IT specialists," *SHS Web of Conferences*, vol. 104, p. 03007, 2021, doi: 10.1051/shsconf/202110403007.
- [2] C. M. Stracke, S. Downes, G. Conole, D. Burgos, and F. Nascimbeni, "Are MOOCs Open Educational Resources? A literature review on history, definitions and typologies of OER and MOOCs," *Open Praxis*, vol. 11, no. 4, p. 331, Dec. 2019, doi: 10.5944/openpraxis.11.4.1010.
- [3] D. Shah, "By The Numbers: MOOCs in 2021," The Report. Accessed: Oct. 05, 2023. [Online]. Available: <https://www.classcentral.com/report/mooc-stats-2021/>
- [4] A. Mcauley, B. Stewart, G. Siemens, and D. Cormier, "THE MOOC MODEL FOR DIGITAL PRACTICE: Digital ways of knowing and learning," 2010. [Online]. Available: <http://www.flickr.com/photos/ecstaticist/3570660643/>
- [5] I. Arpaci, M. Al-Emran, and M. A. Al-Sharafi, "The impact of knowledge management practices on the acceptance of Massive Open Online Courses (MOOCs) by engineering students: A cross-cultural comparison," *Telematics and Informatics*, vol. 54, Nov. 2020, doi: 10.1016/j.tele.2020.101468.
- [6] W. M. Al-Rahmi, N. Yahaya, M. M. Alamri, I. Y. Alyoussef, A. M. Al-Rahmi, and Y. Bin Kamin, "Integrating innovation diffusion theory with technology acceptance model: supporting students' attitude towards using a massive open online courses (MOOCs) systems," *Interactive Learning Environments*, vol. 29, no. 8, pp. 1380–1392, 2021, doi: 10.1080/10494820.2019.1629599.
- [7] Y. Aljaraideh, "Massive Open Online Learning (MOOC) benefits and challenges: A case study in Jordanian context," *International Journal of Instruction*, vol. 12, no. 4, pp. 65–78, Oct. 2019, doi: 10.29333/iji.2019.1245a.
- [8] R. Corrado, E. Pretorius, and G. Van Der Westhuizen, "Undergraduate students' experiences of the use of moocs for learning at a cambodian university," *Educ Sci (Basel)*, vol. 11, no. 7, Jul. 2021, doi: 10.3390/educsci11070336.
- [9] J. S. Brereton and K. Young, "Establishing Social Learning in an Engineering MOOC: Benefits for Diversity and Inclusion in Engineering Education," *Sustainability (Switzerland)*, vol. 14, no. 9, May 2022, doi: 10.3390/su14095472.
- [10] K. R. M. Rafiq, H. Hashim, and M. Md Yunus, "MOOC for Training: How Far It Benefits Employees?," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Dec. 2019. doi: 10.1088/1742-6596/1424/1/012033.
- [11] S. Edelsbrunner, K. Steiner, S. Schön, M. Ebner, and P. Leitner, "Promoting Digital Skills for Austrian Employees through a MOOC: Results and Lessons Learned from Design and Implementation," *Educ Sci (Basel)*, vol. 12, no. 2, Feb. 2022, doi: 10.3390/educsci12020089.
- [12] M. Mehrabi, A. R. Safarpour, and A. A. Keshkar, "Massive Open Online Courses (MOOCs) Dropout Rate in the World: A Protocol for Systematic Review and Meta-analysis," *Interdiscip J Virtual Learn Med Sci*, vol. 13, no. 2, p. 86, 2022, doi: 10.30476/IJVLMS.2022.94572.1138.
- [13] P. Topali, A. Ortega-Arranz, E. Er, A. Martínez-Monés, S. L. Villagrà-Sobrino, and Y. Dimitriadis, "Exploring the problems experienced by learners in a MOOC implementing active learning pedagogies," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2019, pp. 81–90. doi: 10.1007/978-3-030-19875-6\_10.
- [14] P. Topali, A. Ortega-Arranz, A. Martínez-Monés, and S. L. Villagrà-Sobrino,

- “Houston, we have a problem’: Revealing MOOC practitioners’ experiences regarding feedback provision to learners facing difficulties,” *Computer Applications in Engineering Education*, vol. 29, no. 4, pp. 769–785, Jul. 2021, doi: 10.1002/cae.22360.
- [15] C. Chen, G. Sonnert, P. M. Sadler, D. D. Sasselov, C. Fredericks, and D. J. Malan, “Going over the cliff: MOOC dropout behavior at chapter transition,” *Distance Education*, vol. 41, no. 1, pp. 6–25, Jan. 2020, doi: 10.1080/01587919.2020.1724772.
- [16] G. R. El Said, “Understanding How Learners Use Massive Open Online Courses and Why They Drop Out,” *Journal of Educational Computing Research*, vol. 55, no. 5, pp. 724–752, Sep. 2017, doi: 10.1177/0735633116681302.
- [17] M. Garreta-Domingo, P. B. Sloep, and D. Hernández-Leo, “Human-centred design to empower ‘teachers as designers,’” *British Journal of Educational Technology*, vol. 49, no. 6, pp. 1113–1130, Nov. 2018, doi: 10.1111/bjet.12682.
- [18] J. Goopio and C. Cheung, “The MOOC dropout phenomenon and retention strategies,” *Journal of Teaching in Travel and Tourism*, vol. 21, no. 2, pp. 177–197, 2021, doi: 10.1080/15313220.2020.1809050.
- [19] J. Wang, P. Lin, Y. Li, R. Xu, P. Nie, and Y. Xu, “Problems and solutions of MOOC application in provincial colleges and universities,” in *The 14th International Conference on Computer Science & Education (ICCSE 2019)*, 2019, pp. 796–801.
- [20] K. Rabahallah, L. Mahdaoui, and F. Azouaou, “MOOCs recommender system using ontology and memory-based collaborative filtering,” in *ICEIS 2018 - Proceedings of the 20th International Conference on Enterprise Information Systems*, SciTePress, 2018, pp. 635–641. doi: 10.5220/0006786006350641.
- [21] Y. Wang, B. Liang, W. Ji, S. Wang, and Y. Chen, “An improved algorithm for personalized recommendation on MOOCs,” *International Journal of Crowd Science*, vol. 1, no. 3, pp. 186–196, Sep. 2017, doi: 10.1108/IJCS-08-2017-0021.
- [22] I. Fahrurrozi, E. M. D. Admoko, and A. Susilo, “PERBANDINGAN METODE COLLABORATIVE FILTERING DAN HYBRID SEMANTIC SIMILARITY,” *Jurnal Nasional Teknologi Terapan*, vol. 2, no. 3, pp. 334–342, 2018.
- [23] S. Sylvia and S. Lestari, “Implementasi K-Means Dalam Mengatasi Masalah Cold Star Pada Collaborative Filtering,” in *Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat*, 2022, pp. 58–63.
- [24] E. Çano and M. Morisio, “Hybrid recommender systems: A systematic literature review,” 2017, *IOS Press*. doi: 10.3233/IDA-163209.
- [25] K. B. P. Jati, A. T. Wibowo, and R. Rismala, “Implementasi dan Analisis Online-Updating Regularization Kernel Matrix Factorization Model pada Sistem Rekomendasi,” *e-Proceeding of Engineering*, vol. 2, no. 2, pp. 6378–6386, 2015.
- [26] H. Nurrahmi, A. T. Wibowo, and S. Meliana, “Non-Negative Matrix Factorization Based Recommender System using Female Daily Implicit Feedback,” *Ind. Journal on Computing*, vol. 7, no. 1, pp. 1–14, 2022, doi: 10.34818/indojc.2022.7.1.599.
- [27] F. B. A. Larasati and H. Februariyanti, “SISTEM REKOMENDASI PRODUCT EMINA COSMETICS DENGAN MENGGUNAKAN METODE CONTENT - BASED FILTERING,” *MISI (Jurnal Manajemen informatika & Sistem Informasi)*, vol. 4, no. 1, pp. 45–54, 2021.
- [28] A. A. Widjaja and H. N. Palit, “Hybrid Recommendation System untuk Peminjaman Buku Perpustakaan dengan Collaborative dan Content-Based Filtering,” *JURNAL INFRA*, vol. 10, no. 2, pp. 1–6, 2022.
- [29] M. I. Fathurrahman, D. Nurjanah, and R. Rismala, “Sistem Rekomendasi pada Buku dengan Menggunakan Metode Trust-Aware Recommendation Recommendation System for book by using Trust-Aware Recommendation Method,” in *e-Proceeding of Engineering*, 2017, pp. 4966–4977.
- [30] S. Souabi, A. Retbi, M. K. Idrissi, and S. Bennani, “Recommendation Systems on E-Learning and Social Learning: A Systematic Review,” *The Electronic Journal of e-Learning*, vol. 19, no. 5, pp. 432–451, 2021, [Online]. Available: www.ejel.org
- [31] J. Joy and R. V. G. Pillai, “Review and classification of content recommenders in E-learning environment,” 2021, *King Saud bin Abdulaziz University*. doi: 10.1016/j.jksuci.2021.06.009.
- [32] E. Erlangga and H. Sutrisno, “Sistem Rekomendasi Beauty Shop Berbasis Collaborative Filtering,” *Jurnal Manajemen Sistem Informasi dan Teknologi*, vol. 10, no. 2, pp. 47–52, 2020.
- [33] H. Februariyanti, A. Dwi Laksono, J. Sasongko Wibowo, and M. Siswo Utomo, “IMPLEMENTASI METODE COLLABORATIVE FILTERING UNTUK SISTEM REKOMENDASI PENJUALAN

- PADA TOKO MEBEL,” *JURNAL KHATULISTIWA INFORMATIKA*, vol. 9, no. 1, pp. 43–50, 2021, [Online]. Available: [www.unisbank.ac.id](http://www.unisbank.ac.id)
- [34] A. E. Wijaya and D. Alfian, “SISTEM REKOMENDASI LAPTOP MENGGUNAKAN COLLABORATIVE FILTERING DAN CONTENT-BASED FILTERING,” *Jurnal Computech & Bisnis*, vol. 12, no. 1, pp. 11–27, 2018.
- [35] I. M. I. Subroto, S. Mulyono, R. Firmansyah, M. Qomaruddin, and E. N. B. Susila, “Sistem Rekomendasi pada Pembelajaran Mobile Menggunakan Metode Cosine Similarity dan Collaborative Filtering,” *Jurnal Transistor Elektro dan Informatika (TRANSISTOR EI)*, vol. 4, no. 1, pp. 21–28, 2022.
- [36] P. Parhi, A. Pal, and M. Aggarwal, “A Survey of Methods of Collaborative Filtering Techniques,” in *International Conference on Inventive Systems and Control (ICISC-2017)*, 2017, pp. 1–7.
- [37] A. Adawiyah and S. R. Nudin, “Penerapan Algoritma Content Based Filtering dan Frequent Pattern Growth pada Sistem Rekomendasi Program Mahasiswa Wirausaha di Universitas Negeri Surabaya,” *Journal of Informatics and Computer Science*, vol. 3, no. 2, pp. 123–130, 2021.
- [38] M. B. S. Siddik and A. T. Wibowo, “Collaborative Filtering Based Food Recommendation System Using Matrix Factorization,” *Jurnal Media Informatika Budidarma*, vol. 7, no. 3, pp. 1041–1049, 2023, doi: 10.30865/mib.v7i3.6049.
- [39] K. K. Jena *et al.*, “E-Learning Course Recommender System Using Collaborative Filtering Models,” *Electronics (Switzerland)*, vol. 12, no. 1, Jan. 2023, doi: 10.3390/electronics12010157.
- [40] A. N. Khusna, K. P. Delasano, and D. C. E. Saputra, “Penerapan User-Based Collaborative Filtering Algorithm Studi Kasus Sistem Rekomendasi untuk Menentukan Gadget Shield,” *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 2, pp. 293–304, May 2021, doi: 10.30812/matrik.v20i2.1124.
- [41] A. 'Alim, A. Solichin, and Painem, “Rekomendasi Keterampilan Teknologi Informasi Menggunakan Metode User-Based Collaborative Filtering dan Log-Likelihood Similarity,” *Cogito Smart Journal*, vol. 6, no. 2, pp. 141–154, 2020.
- [42] M. Zarlis, R. Widia Sembiring, S. Tunas Bangsa Pematangsiantar, and J. A. Jend Sudirman Blok No, “ANALISA TERHADAP PERBANDINGAN ALGORITMA DECISION TREE DENGAN ALGORITMA RANDOM TREE UNTUK PRE-PROCESSING DATA,” *Jurnal Sains Komputer & Informatika (J-SAKTI)*, no. 1, 2017, [Online]. Available: <http://tunasbangsa.ac.id/ejurnal/index.php/jsakti>
- [43] N. Sagala and H. Tampubolon, “Kinerja Algoritma Data Mining pada Dataset Konsumsi Alkohol Siswa,” *Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 2, pp. 98–103, 2018.
- [44] E. Etriyanti, D. Syamsuar, and Y. N. Kunang, “Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa,” *Telematika*, vol. 13, no. 1, pp. 56–67, Feb. 2020, doi: 10.35671/telematika.v13i1.881.
- [45] A. Izzah and R. Widyastuti, “Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression untuk Pencegahan Data Outlier,” *KINETIK*, vol. 2, no. 3, pp. 141–150, Jul. 2017, doi: 10.22219/kinetik.v2i3.268.
- [46] E. D. Wahyuni, A. A. Arifiyanti, and M. Kustyani, “Exploratory Data Analysis dalam Konteks Klasifikasi Data Mining,” in *Prosiding Nasional Rekayasa Teknologi Industri dan Informasi XIV Tahun 2019 (ReTII)*, 2019, pp. 263–269. [Online]. Available: <http://journal.itny.ac.id/index.php/ReTII>
- [47] H. Sulastri and A. I. Gufroni, “PENERAPAN DATA MINING DALAM PENGELOMPOKAN PENDERITA THALASSAEMIA,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 3, no. 2, pp. 299–305, Sep. 2017, doi: 10.25077/teknosi.v3i2.2017.299-305.
- [48] V. R. Prasetyo, M. Mercifia, A. Averina, L. Sunyoto, and Budiarmo, “PREDIKSI RATING FILM PADA WEBSITE IMDB MENGGUNAKAN METODE NEURAL NETWORK,” *NERO: Networking Engineering Research Operation*, vol. 7, no. 1, pp. 1–8, 2022.
- [49] F. Astuti, R. M. Candra, S. Agustian, and S. Ramadhani, “Klasifikasi Sentimen Masyarakat Terhadap Pemerintah Terkait Penerapan Kebijakan New Normal Menggunakan Metode K-Nearest Neighbor,” *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 3, pp. 531–538, 2022.
- [50] I. Yoshua and H. Bunyamin, “Pengimplementasian Sistem Rekomendasi Musik Dengan Metode Collaborative Filtering,” *Jurnal Strategi*, vol. 3, no. 1, pp.

1–16, 2021.

- [51] M. H. Aghdam, M. Analoui, and P. Kabiri, “Collaborative filtering using non-negative matrix factorisation,” *J Inf Sci*, vol. 43, no. 4, pp. 567–579, Aug. 2017, doi: 10.1177/0165551516654354.
- [52] G. M. Del Corso and F. Romani, “Adaptive nonnegative matrix factorization and measure comparisons for recommender systems,” *Appl Math Comput*, vol. 354, pp. 164–179, Aug. 2019, doi: 10.1016/j.amc.2019.01.047.
- [53] Z. Yang, W. Chen, and J. Huang, “Enhancing recommendation on extremely sparse data with blocks-coupled non-negative matrix factorization,” *Neurocomputing*, vol. 278, pp. 126–133, Feb. 2018, doi: 10.1016/j.neucom.2017.04.080.
- [54] H. H. Arfisko and A. T. Wibowo, “Sistem Rekomendasi Film Menggunakan Metode Hybrid Collaborative Filtering Dan Content-Based Filtering,” *e-Proceeding of Engineering*, vol. 9, no. 3, pp. 2149–2159, 2022.
- [55] M. E. B. H. Kbaier, H. Masri, and S. Krichen, “A personalized hybrid tourism recommender system,” in *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, IEEE Computer Society, Jul. 2017, pp. 244–250. doi: 10.1109/AICCSA.2017.12.
- [56] S. Suriati, M. Dwiastuti, and T. Tulus, “Weighted hybrid technique for recommender system,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Dec. 2017. doi: 10.1088/1742-6596/930/1/012050.