

RANCANG BANGUN SISTEM REKOMENDASI BUKU BERBASIS ITEM-BASED COLLABORATIVE FILTERING MENGUNAKAN ALGORITMA K-NEAREST NEIGHBORS

Vernanda Sam Saputra ^{a,1,*}, Ahmad Ridwan ^{b,2}, Taftazani Ghazi Pratama ^{c,3}

^{a,b,c} Universitas Muhammadiyah Kudus, Jl. Ganesha Raya No.1, Purwosari, Kabupaten Kudus, Jawa Tengah, Indonesia

¹ vernandasam@gmail.com*; ² achmadridwan@umkudus.ac.id; ³ taftazanighazi@umkudus.ac.id

* corresponding author

ARTICLE INFO

Keywords:

K Nearests Neighbors, Machine Learning, Recommendation System

ABSTRACT

This research aims to design and develop a book recommendation system using the item-based collaborative filtering method with the K-Nearest Neighbors algorithm. The K-Nearest Neighbors algorithm was chosen because of its ability to measure similarity between items based on distance calculations, thereby providing relevant recommendations for users. The method involves processing user rating data for books, followed by the development of a model using an item-based approach to find similar books. The model is tested using cross-validation techniques with evaluation metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) to measure the accuracy of the recommendations. The test results indicate that the proposed model has a good level of accuracy, with an average RMSE value of 0.8191 and an MAE of 0.6235. This system is expected to enhance the user experience in finding books that match their preferences and to contribute significantly to the development of machine learning-based recommendation systems.

1. Pendahuluan

Di era digital saat ini, teknologi informasi mengalami kemajuan yang sangat cepat, dan pengaruh teknologi pun mulai merambah ke hampir semua aspek kehidupan manusia [1]. Dalam era digital yang semakin berkembang, informasi dan pilihan yang tersedia bagi pengguna menjadi semakin melimpah [2]. Hal ini menimbulkan tantangan dalam menemukan informasi atau produk yang sesuai dengan preferensi dan kebutuhan individu. Salah satu pendekatan yang efektif untuk menangani masalah ini adalah dengan menggunakan sistem rekomendasi. Sistem Rekomendasi merupakan model aplikasi yang dapat membantu pengguna untuk mengambil keputusan tentang produk yang akan dipilih berdasarkan observasi atas keinginan mereka [3]. Sistem rekomendasi bertujuan untuk menyaring informasi secara efisien dan memberikan rekomendasi yang relevan bagi pengguna. Pada sektor literatur, seperti platform buku digital dan e-commerce, sistem rekomendasi telah menjadi komponen penting dalam meningkatkan pengalaman pengguna dengan menyarankan buku-buku yang sesuai berdasarkan preferensi pengguna sebelumnya.

Sistem rekomendasi memiliki banyak metode, salah satu metode yang sering digunakan untuk mencari rekomendasi adalah collaborative filtering [4]. Pendekatan collaborative filtering telah menjadi salah satu metode populer dalam pengembangan sistem rekomendasi. Metode ini bekerja dengan menganalisis pola perilaku pengguna, seperti ulasan atau penilaian terhadap item tertentu, untuk menemukan kesamaan antara item atau pengguna. Di antara dua pendekatan utama dalam collaborative filtering user-based dan item-based penelitian ini memfokuskan pada pendekatan item-based collaborative filtering, yang menekankan pada

analisis kesamaan antar item berdasarkan perilaku pengguna. Collaborative filtering memberikan rekomendasi berdasarkan kumpulan dari pendapat, minat dan ketertarikan beberapa pengguna yang biasanya diberikan dalam bentuk rating yang diberikan pengguna kepada suatu item [5]. Pendekatan ini dinilai lebih stabil dan scalable dibandingkan pendekatan user-based, terutama ketika jumlah pengguna jauh lebih besar daripada jumlah item, seperti pada konteks rekomendasi buku. Penerapan metode item-based collaborative filtering juga menghasilkan rekomendasi yang lebih cepat [6]. Penerapan metode item-based collaborative filtering menggunakan lebih sedikit memori dan waktu dalam menghitung nilai kemiripan antar buku, penerapan metode item-based collaborative filtering juga lebih baik digunakan untuk data yang cenderung statis [7].

Algoritma K-Nearest Neighbors (K-NN) dipilih sebagai metode untuk mengimplementasikan item-based collaborative filtering dalam penelitian ini. K-NN adalah algoritma yang sering digunakan dalam tugas-tugas klasifikasi dan regresi, termasuk dalam sistem rekomendasi, karena kemampuannya yang sederhana namun efektif dalam menghitung jarak atau kesamaan antara data. Algoritma K-Nearest-Neighbor adalah sebuah metode yang bisa digunakan untuk klasifikasi objek berdasarkan atribut dan sampel latih yang telah tersedia [8]. Pada konteks sistem rekomendasi, algoritma K-NN digunakan untuk mengidentifikasi buku-buku yang memiliki kesamaan karakteristik berdasarkan penilaian pengguna terhadap buku-buku tersebut. Setelah kesamaan dihitung, sistem dapat memberikan rekomendasi buku yang paling relevan untuk setiap pengguna.

Penelitian ini menggunakan dataset yang terdiri dari rating pengguna terhadap berbagai judul buku. Untuk mengevaluasi performa model, dilakukan validasi menggunakan teknik k-fold cross-validation dengan metrik evaluasi Root Mean Square Error (RMSE) dan Mean Absolute Error (MAE). RMSE dan MAE digunakan untuk mengukur seberapa jauh prediksi sistem berbeda dari rating yang sebenarnya, sehingga dapat diketahui tingkat akurasi dari sistem rekomendasi yang dibangun. Sistem rekomendasi dapat digunakan untuk memprediksi barang tertentu yang disukai oleh pengguna atau untuk mengidentifikasi beberapa barang yang mungkin disukai oleh pengguna tertentu [9].

1.1. Sistem Rekomendasi

Sistem rekomendasi adalah sebuah teknik atau metode yang dirancang untuk memberikan saran atau rekomendasi terhadap suatu produk, layanan, atau konten kepada pengguna berdasarkan preferensi dan perilaku mereka. Sistem adalah komponen yang saling berinteraksi, terstruktur dan terorganisasi untuk mencapai tujuan yang telah ditetapkan [10]. Tujuan utama dari sistem rekomendasi adalah untuk membantu pengguna menemukan item yang relevan atau menarik di antara sekumpulan besar pilihan yang tersedia. Sistem ini umumnya digunakan dalam berbagai domain, seperti e-commerce, platform streaming, jejaring sosial, dan perpustakaan digital, untuk meningkatkan pengalaman pengguna dan meningkatkan interaksi dengan layanan tersebut.

Secara umum, ada tiga pendekatan utama dalam sistem rekomendasi, yaitu content-based filtering, collaborative filtering, dan hybrid methods. Content-based filtering memanfaatkan informasi atau atribut dari item yang direkomendasikan dan menyesuaikannya dengan preferensi pengguna berdasarkan karakteristik konten. Sebaliknya, collaborative filtering mengandalkan data dari interaksi pengguna, seperti penilaian, klik, atau perilaku pembelian, untuk menemukan kesamaan di antara pengguna atau item. Pendekatan ini terbagi menjadi dua tipe: user-based dan item-based. Dalam item-based collaborative filtering, sistem menganalisis hubungan antara item berdasarkan pola penilaian oleh pengguna untuk mengidentifikasi item yang mirip dan memberikan rekomendasi.

Keuntungan utama dari sistem rekomendasi adalah kemampuannya untuk mempersonalisasi pengalaman pengguna secara otomatis, meningkatkan keterlibatan dan loyalitas pengguna, serta mengoptimalkan pengambilan keputusan. Namun, tantangan seperti sparsity data, cold start problem, dan kualitas prediksi

tetap menjadi fokus utama dalam penelitian dan pengembangan di bidang ini. Oleh karena itu, berbagai teknik dan algoritma, seperti K-Nearest Neighbors (K-NN), digunakan untuk meningkatkan akurasi dan relevansi rekomendasi yang diberikan kepada pengguna.

1.2. K-Nearests Neighbors

Nearest Neighbors (K-NN) adalah salah satu algoritma pembelajaran mesin yang paling sederhana dan efektif yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini berbasis pada prinsip "kedekatan" atau "kemiripan" antar data, dengan cara mengidentifikasi sejumlah k tetangga terdekat dari sebuah titik data yang ingin diklasifikasikan atau diprediksi. Titik data baru diklasifikasikan berdasarkan mayoritas kelas dari tetangganya atau diberi nilai rata-rata tetangganya jika digunakan untuk regresi.

Pada konteks sistem rekomendasi, khususnya item-based collaborative filtering, algoritma K-NN digunakan untuk menemukan item-item yang memiliki kemiripan dengan item yang telah dinilai oleh pengguna. Kemiripan antar item dihitung berdasarkan pola penilaian dari pengguna lain terhadap kedua item tersebut. Metode ini memungkinkan sistem untuk memberikan rekomendasi item yang dianggap serupa dengan preferensi pengguna berdasarkan interaksi sebelumnya. Kemiripan antar item sering kali diukur menggunakan berbagai metrik seperti cosine similarity, Pearson correlation, atau Euclidean distance. Setelah kemiripan dihitung, algoritma K-NN akan mencari k item terdekat yang paling mirip dengan item target. Nilai k ini dapat disesuaikan berdasarkan kebutuhan dan kompleksitas data yang dianalisis.

Algoritma K-NN memiliki keuntungan karena mudah diimplementasikan dan interpretatif, serta bekerja baik pada dataset yang tidak memiliki asumsi distribusi tertentu. Namun, algoritma ini memiliki beberapa kelemahan, seperti sensitivitas terhadap dimensi data yang tinggi dan memerlukan waktu komputasi yang besar jika dataset yang digunakan sangat besar. Oleh karena itu, pemilihan dan pengaturan parameter yang tepat, seperti nilai k , sangat penting untuk mengoptimalkan kinerja algoritma dalam konteks tertentu.

1.3. Collaborative Filtering

Collaborative Filtering (CF) adalah salah satu teknik utama dalam sistem rekomendasi yang memanfaatkan informasi dan preferensi pengguna lain untuk menghasilkan rekomendasi. CF bekerja berdasarkan prinsip bahwa pengguna yang memiliki preferensi serupa di masa lalu kemungkinan besar akan menyukai hal yang sama di masa depan. Dalam penerapannya, CF dapat dibagi menjadi dua kategori utama: user-based collaborative filtering dan item-based collaborative filtering.

Pada user-based collaborative filtering, rekomendasi diberikan kepada pengguna berdasarkan kesamaan preferensi dengan pengguna lain. Sedangkan pada item-based collaborative filtering, algoritma fokus pada kesamaan antar item. Sistem ini menganalisis pola penilaian yang diberikan pengguna pada item tertentu untuk menemukan item-item lain yang serupa. Item yang serupa dianggap memiliki pola penilaian yang sama dari banyak pengguna.

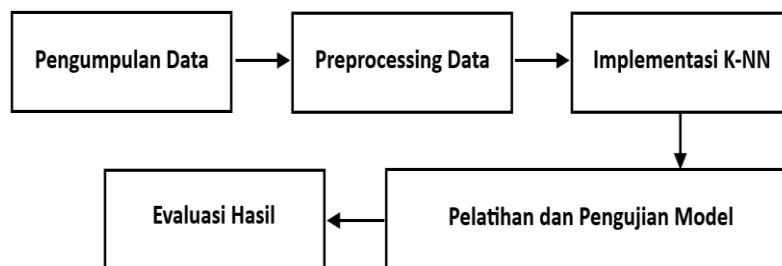
Dalam konteks item-based collaborative filtering, langkah pertama adalah menghitung kemiripan antar item berdasarkan preferensi pengguna. Kemiripan ini bisa dihitung menggunakan berbagai metrik seperti cosine similarity, Pearson correlation coefficient, atau adjusted cosine similarity. Setelah kemiripan antar item dihitung, rekomendasi diberikan dengan memilih item yang paling mirip dengan item yang telah dinilai positif oleh pengguna.

Pendekatan ini memiliki beberapa keunggulan, seperti keakuratan yang lebih tinggi dalam rekomendasi dibandingkan dengan pendekatan berbasis pengguna karena item cenderung memiliki lebih banyak interaksi, serta skalabilitas yang lebih baik pada dataset besar. Namun, pendekatan ini juga memiliki tantangan, seperti membutuhkan waktu komputasi yang signifikan untuk menghitung kemiripan antar item, terutama pada dataset yang sangat besar

Dengan memanfaatkan item-based collaborative filtering, sistem rekomendasi dapat lebih efektif dan relevan dalam memberikan rekomendasi buku yang sesuai dengan preferensi pengguna berdasarkan pola penilaian terhadap buku-buku serupa yang telah diberikan oleh pengguna lain.

2. Metodologi Penelitian

Metode penelitian ini fokus pada pengembangan model sistem rekomendasi yang dievaluasi menggunakan metrik *Root Mean Square Error* (RMSE) dan *Mean Absolute Error* (MAE) untuk mengukur akurasi prediksi. Dalam penelitian ini, Proses penelitian dilakukan secara bertahap, tahapan-tahapan penelitian ini meliputi dengan pengumpulan data, prapemrosesan data, implementasi algoritma K-NN, pengujian dan evaluasi model, serta analisis hasil berdasarkan kinerja sistem. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Tahapan pertama dalam penelitian ini adalah pengumpulan data. Dataset yang digunakan adalah data penilaian (rating) buku yang diperoleh dari platform Goodreads, yang mencakup sekitar 2.395.300 penilaian yang diberikan oleh pengguna terhadap berbagai buku. Dataset ini mengandung beberapa atribut penting seperti `user_id`, `book_id`, dan `rating`, yang merepresentasikan ID pengguna, ID buku, dan skor penilaian pengguna terhadap buku, masing-masing. Selain itu, data buku tambahan yang mencakup atribut seperti `title`, `authors`, dan `publication_year` juga digunakan untuk mengidentifikasi informasi lebih lanjut tentang buku yang direkomendasikan.

2.2. Preprocessing Data

Prapemrosesan data merupakan langkah penting untuk memastikan data siap digunakan dalam model pembelajaran mesin. Langkah-langkah pra-pemrosesan yang dilakukan dalam penelitian ini yaitu pembersihan data yang dilakukan untuk menghilangkan atribut yang tidak relevan atau redundan. Kemudian, dataset diperiksa untuk mengidentifikasi nilai yang hilang dan dilakukan penanganan nilai hilang yang kemudian diisi dengan rata-rata penilaian yang ada untuk memastikan konsistensi data. Lalu, Transformasi Data yang mana data disiapkan dalam format yang sesuai untuk digunakan. Selain itu, data juga diubah ke format dataframe yang digunakan untuk proses pelatihan dan pengujian.

2.3. Implementasi K-NN

Pada tahap implementasi algoritma K-NN, algoritma K-NN diimplementasikan untuk mengembangkan sistem rekomendasi buku berbasis item-based collaborative filtering. Dilakukan pembagian data, dataset dibagi menjadi dua bagian yaitu data pelatihan dan data pengujian. Sebanyak 75% data digunakan untuk pelatihan model, dan 25% sisanya digunakan untuk pengujian. Algoritma K-NN yang digunakan adalah *KNN Baseline*, yang merupakan salah satu varian K-NN dengan pilihan *similarity metric* seperti *cosine similarity*

dan *pearson correlation*. Metode KNN Baseline dipilih karena mampu menangani data skala besar dan mengoptimalkan kesamaan antar item dengan menggunakan model baseline.

2.4. Pengujian dan Evaluasi Model

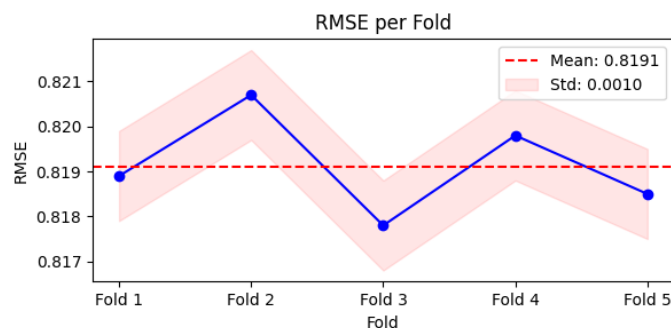
Pada tahap pengujian dan evaluasi model, model dilatih menggunakan data pelatihan dan kesamaan antar buku dihitung dengan menggunakan cosine similarity sebagai ukuran jarak. Setelah model dilatih, model diuji dengan data pengujian untuk menghasilkan prediksi penilaian buku oleh pengguna. Metrik evaluasi yang digunakan adalah *Root Mean Square Error* (RMSE) dan *Mean Absolute Error* (MAE) untuk mengukur akurasi prediksi.

3. Hasil dan Pembahasan

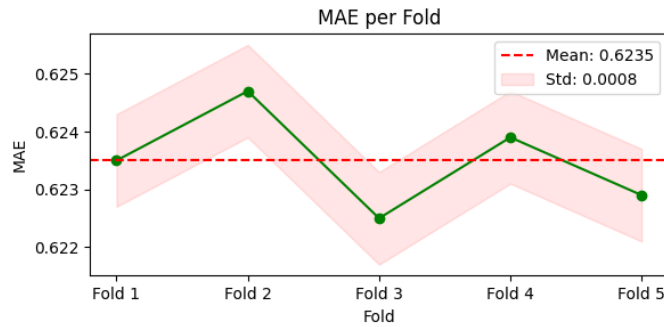
3.1. Hasil dan Evaluasi Model

Setelah melalui tahapan implementasi, model sistem rekomendasi diuji menggunakan teknik cross-validation sebanyak 5 kali (5-fold cross-validation). Penggunaan cross-validation bertujuan untuk mengukur kinerja model dalam menghasilkan rekomendasi yang akurat pada data yang tidak pernah dilihat sebelumnya. Metrik evaluasi yang digunakan adalah *Root Mean Square Error* (RMSE) dan *Mean Absolute Error* (MAE).

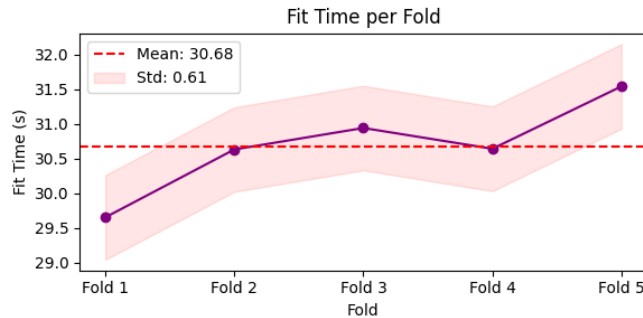
Hasil evaluasi dari model menunjukkan nilai RMSE rata-rata sebesar 0.819 dan nilai MAE rata-rata sebesar 0.624. Nilai RMSE dan MAE yang relatif rendah ini menunjukkan bahwa sistem rekomendasi yang dikembangkan mampu memprediksi rating buku dengan cukup akurat, mendekati nilai rating yang sebenarnya diberikan oleh pengguna. Nilai RMSE yang lebih rendah dibandingkan MAE juga menunjukkan bahwa sistem lebih sensitif terhadap kesalahan prediksi yang besar, dan performa keseluruhannya konsisten di seluruh fold pengujian. Hasil evaluasi model dapat dilihat pada Gambar 2, 3, 4, dan 5.



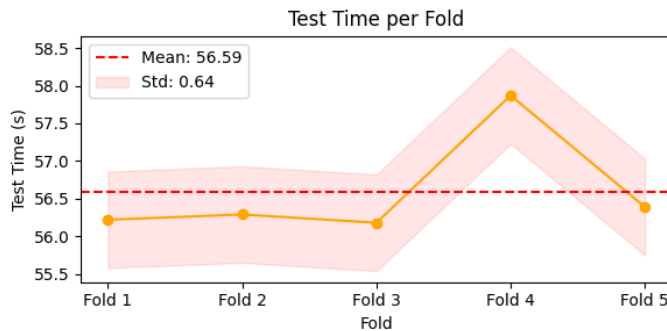
Gambar 2. Hasil Testset RMSE



Gambar 3. Hasil Testset MAE



Gambar 4. Hasil Fit Time 5 Fold



Gambar 5. Hasil Test Time 5 Fold

3.2. Pengujian

Dilakukan juga untuk mengidentifikasi buku-buku yang direkomendasikan oleh sistem berdasarkan input buku tertentu. Misalnya, untuk buku berjudul "How to Win Friends and Influence People" oleh Dale Carnegie, Sistem merekomendasikan sepuluh buku lain dengan jarak kesamaan (similarity distance) yang tinggi, yaitu nilai cosine similarity yang kurang lebih sama, yaitu sebesar 0.7 untuk setiap buku.

Buku-buku yang direkomendasikan mencakup berbagai genre dan penulis, seperti "The House of God" oleh Samuel Shem, "Holy Cow: An Indian Adventure" oleh Sarah Macdonald, "Housekeeping" oleh Marilynne Robinson, dan lain-lain. Kesamaan jarak yang seragam ini mengindikasikan bahwa buku-buku tersebut memiliki atribut penilaian yang sangat mirip oleh pengguna, yang menjadi dasar untuk rekomendasi.

4. Kesimpulan



Berdasarkan penelitian yang telah dilakukan mengenai rancang bangun sistem rekomendasi buku, kesimpulan yang dapat diambil sesuai dengan tujuan yang telah disebutkan di awal, adalah sebagai berikut:

1. Algoritma K-Nearest Neighbors dapat diimplementasikan pada sistem rekomendasi buku berbasis item-based collaborative filtering.
2. Algoritma k-NN mampu memberikan rekomendasi dengan akurasi tinggi pada dataset besar dengan optimalisasi parameter, seperti nilai k dan normalisasi data.

Hasil penelitian menunjukkan bahwa algoritma k-NN cukup efektif untuk membangun sistem rekomendasi berbasis data rating pengguna. Sistem ini berhasil mengatasi sebagian besar kendala teknis melalui pendekatan optimasi, namun terdapat keterbatasan, terutama dalam menangani masalah *cold-start* dan kurangnya informasi tambahan. Dengan demikian, penelitian ini memberikan dasar yang kuat untuk pengembangan sistem rekomendasi yang lebih kompleks di masa mendatang.

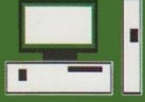
Berdasarkan hasil penelitian, terdapat beberapa saran untuk pengembangan dan implementasi sistem rekomendasi buku sebagai berikut:

1. Pengembangan Sistem Hybrid Filtering. Dengan mengintegrasikan algoritma content-based filtering dengan collaborative filtering agar sistem dapat memanfaatkan metadata buku, seperti genre, penulis, atau tahun publikasi.
2. Optimalisasi Data Sparsity. Menerapkan teknik pengurangan dimensi, seperti matrix factorization, untuk menangani sparsitas data yang sering menjadi hambatan dalam rekomendasi berbasis K-NN.
3. Penggunaan Data Kontekstual. Memasukkan data demografi pengguna, seperti usia, lokasi, dan preferensi membaca, untuk memberikan rekomendasi yang lebih personal.

Saran-saran ini diharapkan dapat menjadi panduan bagi pengembang dan peneliti selanjutnya untuk meningkatkan kualitas sistem rekomendasi berbasis K-NN, sekaligus memberikan solusi yang lebih komprehensif dalam mengatasi tantangan yang ada. Dengan mengimplementasikan langkah-langkah tersebut, sistem rekomendasi buku dapat menjadi lebih relevan, akurat, dan efektif dalam memenuhi kebutuhan pengguna

Daftar Pustaka

- [1] H. Ardiansyah and M. A. Bianto, "Implementation of License Plate Recognition Monitoring System using Neural Network on Solar Powered Microcontroller," *Indonesian Vocational Research Journal*, vol. 2, no. 1. p. 105, 2022, doi: 10.30587/ivrj.v2i1.4949.
- [2] E. Handoyo, "Analisis Tingkat Keamanan Informasi: Studi Komparasi Framework Cobit 5 Subdomain Manage Security Services (DSS05) dan NIST Sp 800 – 55," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 1, no. 2. pp. 76–83, 2020, doi: 10.37859/coscitech.v1i2.2199.
- [3] E. Erlangga and H. Sutrisno, "Sistem Rekomendasi Beauty Shop Berbasis Collaborative Filtering," *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, vol. 10, no. 2. p. 47, 2020, doi: 10.36448/jmsit.v10i2.1611.
- [4] V. L. Jaja, B. Susanto, and L. R. Sasongko, "Penerapan Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Data MovieLens," *d'CARTESIAN*, vol. 9, no. 2. p. 78, 2020, doi: 10.35799/dc.9.2.2020.28274.
- [5] T. Badriyah, I. Restuningtyas, and F. Setyorini, "Sistem Rekomendasi Collaborative Filtering Berbasis User Algoritma Adjusted Cosine Similarity," *Prosiding Seminar Nasional Sisfotek*, vol. 10, no. 1. pp. 38–45, 2021.
- [6] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," *Proceedings of the 10th International Conference on World Wide Web, WWW 2001*. pp. 285–295, 2001, doi: 10.1145/371920.372071.
- [7] F. Ricci, L. Rokach, B. Shapira, and P. Kantor, "Recommender Systems Handbook," *Recommender Systems*



Handbook. 2022, doi: 10.1007/978-1-0716-2197-4.

- [8] Yahya and W. Puspita Hidayanti, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada ‘Lombok Vape On,’” *Infotek : Jurnal Informatika dan Teknologi*, vol. 3, no. 2. pp. 104–114, 2020, doi: 10.29408/jit.v3i2.2279.
- [9] Deshpande, Mukund and Karypis, George, “Item-based Top-N Recommendation Algorithms,” *ACM Transactions on Information Systems*, vol. 22, no. 1. pp. 147–177, 2004, [Online]. Available: <http://dl.acm.org/citation.cfm?id=963776>.
- [10] O. Triandy and Tukino, “Sistem Pendukung Keputusan Penerimaan Karyawan Baru Pada PT Kingslee Infinitas Teknologi,” *Jurnal Comasie*, vol. 4, no. 3. pp. 57–66, 2021.