

Perbandingan *Cosine Similarity* Dan *Euclidean Distance* Pada Model Rekomendasi Buku Dengan Metode *Item-Based Collaborative Filtering*

M. Dzikri Hisyam Ilyasa¹, Yuni Yamasari²

^{1,2}Jurusan Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya

dzikri.18023@mhs.unesa.ac.id

yuniyamasari@unesa.ac.id

Abstrak— Model rekomendasi adalah metode penyaringan data atau informasi dengan menggunakan teknik analisis data untuk membantu pengguna dalam menemukan item yang diinginkan. Salah satu metode yang digunakan dalam model rekomendasi adalah *item-based collaborative filtering*. Metode ini merupakan teknik untuk menghitung nilai antara kesamaan item. Algoritma yang dapat digunakan untuk menghitung nilai kesamaan item pada penelitian ini adalah *Cosine Similarity* dan *Euclidean Distance*. Masalahnya adalah masih belum jelas algoritma kesamaan item mana yang lebih akurat antara model rekomendasi yang menggunakan *Cosine Similarity* atau *Euclidean Distance*. Tujuan dari penelitian ini adalah untuk mengetahui algoritma kesamaan item mana yang paling akurat antara *Cosine Similarity* dan *Euclidean Distance*. Dari hasil pengujian *Mean Absolute Error (MAE)* menggunakan *K-Fold Cross Validation* pada penelitian ini dapat disimpulkan bahwa rumus perhitungan *Cosine Similarity* mempunyai tingkat akurasi yang cukup baik yaitu 0.647352 dibandingkan dengan rumus perhitungan *Euclidean Distance* memiliki nilai akurasi yaitu 0.676872 dengan skala MAE 0–1. Semakin kecil nilai MAE semakin tinggi akurasi prediksi yang dihasilkan. Hal ini menunjukkan bahwa rumus perhitungan *Cosine Similarity* dapat menghasilkan rekomendasi yang lebih baik dari *Euclidean Distance*.

Kata Kunci— Model Rekomendasi, *Item-based Collaborative Filtering*, *Cosine Similarity*, *Euclidean Distance*, *Mean Absolute Error*.

I. PENDAHULUAN

Perpustakaan merupakan sebuah sarana dimana kita dapat menemukan kumpulan bahan-bahan tertulis, tercetak maupun grafis lainnya seperti film, slide, piringan hitam, tape, dengan berbagai macam jenis informasi yang diatur secara modelatis yang dapat digunakan dalam keperluan studi, penelitian, pembacaan dan lain sebagainya. Di dalam perpustakaan kita dapat mencari buku maupun bahan tertulis lainnya mengenai banyak sekali informasi yang bersifat ilmu pengetahuan, motivasi, kamus, jurnal, rohani, hingga hiburan. Perpustakaan memiliki peranan yang penting dalam pengembangan ilmu pengetahuan dan pendidikan terhadap publik. Menurut (Alharthi, H., Inkpen, D., Szpakowicz, S., 2017), perpustakaan merupakan sarana untuk membantu publik dalam mendukung kemajuan pada berbagai ilmu pengetahuan. Salah satu fungsi dari perpustakaan bukan hanya sebagai repositori sumber referensi semata, namun juga merupakan sarana dalam meningkatkan minat baca masyarakat, yang saat ini menurun pada kalangan anak muda. Model rekomendasi adalah model yang dapat menyaring informasi dalam memprediksi peringkat

atau memberikan preferensi akan suatu item yang diberikan kepada pengguna (E. Uko Okon, B. O. Eke, P. O. Asagba, 2018). Menurut (Hikmatyar, M., Ruuhwan, Zery H., 2019), tujuan di buatnya model rekomendasi buku perpustakaan adalah untuk mempermudah pengunjung dalam mendapatkan referensi yang tepat dan sesuai dengan kebutuhan pengunjung. Model rekomendasi banyak digunakan untuk membuat prediksi seperti buku, musik, film, tempat wisata dan lain-lain. Model rekomendasi yang tepat diperlukan agar rekomendasi yang dibuat dapat sesuai dengan keinginan pengguna dan memudahkan pengunjung dalam mengambil keputusan yang tepat untuk menentukan buku mana yang akan dipilih. Salah satu metode yang dapat digunakan dalam model rekomendasi adalah *Collaborative Filtering*. (Laja Jaja, Y. V., Bambang, S., Leopoldus, R. S., 2020).

Konsep dari *Collaborative Filtering* adalah di mana pendapat pengguna lain akan diperhitungkan untuk memprediksi elemen yang mungkin diminati atau dibutuhkan oleh pengguna. (Wijaya, A. E., Deni A., 2018). Menurut (H. Koohi, K. Kiani, 2017), teknik rekomendasi ini mempunyai tingkat keberhasilan yang tinggi, dan memiliki prospek untuk aplikasi yang luas di berbagai bidang seperti e-commerce dan jejaring sosial. Pada *Collaborative Filtering*, item direkomendasikan kepada pengguna target melalui analisis peringkat pengguna yang paling mirip (*neighbor users*) pada item tersebut. Menurut (Sarwar et al., 2001), metode *collaborative filtering* dibagi menjadi beberapa metode, dua diantaranya adalah *item-based* dan *user-based*. perhitungan *user-based* berjalan dengan memanfaatkan histori pilihan pengguna. Sebagai contoh user A menyukai item X, sedangkan profil pilihan A mirip dengan user B, dan user B belum pernah memberikan penilaian terhadap item X, maka item X akan direkomendasikan kepada user B. Kelemahan dari *user-based* yaitu seiring bertambahnya data user dan item maka bertambah pula kompleksitas perhitungannya. Berbeda dengan metode *user-based*, metode *item-based* tidak menggunakan profil pengguna, melainkan rekomendasi kepada user dihitung dengan menentukan kemiripan antara satu item dengan item lain yang disukai oleh user tersebut.

Item-based collaborative filtering menghitung hubungan satu item ke item lainnya berdasarkan item yang dipilih oleh pengguna, lalu merekomendasikan item tersebut ke pengguna lain berdasarkan korelasi antara item tersebut. Metode ini didasarkan pada kesamaan rating antara suatu produk (Halim,

2013). Pada metode Item-Based Collaborative Filtering, model menghitung kesamaan (*similarity*) antara satu item dan item lainnya yang berdasarkan oleh peringkat yang ditentukan pengguna untuk suatu item.

Penggunaan Collaborative Filtering untuk model rekomendasi dapat menggunakan perhitungan nilai kesamaan *Euclidean Distance* seperti penelitian yang dilakukan oleh (Sari, S., & Hendra, D. T. , 2015). Dari penelitian tersebut disajikan hasil analisis bahwa algoritma perhitungan *Euclidean Distance* dapat memberikan rekomendasi bagi pelanggan pada toko film online dengan nilai Correlation Coefficient yang menunjukkan pendekatan ini memberikan rekomendasi yang cukup mirip dengan keinginan pelanggan. Selain itu, hasil analisis dari penelitian tersebut memiliki hasil bahwa penggunaan Collaborative Filtering dengan *Euclidean Distance* sebagai rumus perhitungan kesamaan nya memiliki tingkat akurasi yang baik.

Terdapat juga penelitian terkait model rekomendasi menggunakan metode Collaborative Filtering dengan *Cosine Similarity* sebagai rumus perhitungan nilai kesamaan seperti penelitian yang dilakukan oleh (Gupta, M., Thakkar, A., Gupta, V., & Rathore, D. P. S. , 2020). Pada penelitian tersebut membandingkan tingkat akurasi dari beberapa pendekatan pada metode Collaborative Filtering pada model rekomendasi menggunakan rumus perhitungan *Cosine Similarity*. Hasilnya tingkat akurasi model rekomendasi yang menggunakan Item-Based Collaborative Filtering memiliki tingkat akurasi yang lebih akurat.

Dengan menerapkan Collaborative Filtering berbasis Item menunjukkan bahwa rekomendasi yang dihasilkan lebih berkualitas dan mendekati dengan kebutuhan pengguna. Selain itu, metode Collaborative Item Filter juga memiliki kelebihan karena dianggap mampu mengungkap tautan implisit, khususnya tautan yang beberapa di antaranya dapat dihasilkan dari riwayat pilihan pengguna, digunakan untuk meningkatkan akurasi rekomendasi yang dihasilkan. (Mustofa, 2018).

Berdasarkan latar belakang diatas, penulis membandingkan tingkat akurasi perhitungan *Cosine Similarity* dan *Euclidean Distance* pada model rekomendasi buku perpustakaan menggunakan metode Item-Based Collaborative Filtering. Penelitian ini bertujuan untuk menguji apakah algoritma perhitungan tingkat kemiripan *Cosine Similarity* lebih baik daripada *Euclidean Distance* atau sebaliknya.

A. Rumusan Masalah

Berdasarkan penjelasan latar belakang diatas, maka rumusan masalah yang akan diambil oleh penulis adalah sebagai berikut:

1. Bagaimana perbandingan tingkat akurasi perhitungan *Cosine Similarity* dan *Euclidean Distance* menggunakan metode item-based collaborative filtering pada model rekomendasi
2. Bagaimana hasil analisis pengujian *Mean Absolute Error* (MAE) menggunakan *K-fold Cross Validation*

pada perhitungan *Cosine Similarity* dan *Euclidean Distance* dengan metode item-based collaborative filtering?

B. Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Hanya tingkat akurasi algoritma *Cosine Similarity* dan *Euclidean Distance* yang dibandingkan dalam penelitian.
2. Hasil pada penelitian berupa nilai similarity, prediksi, tingkat akurasi, dan analisis.
3. Penelitian ini tidak membahas masalah perancangan model rekomendasi, transaksi dan keamanan data pada perpustakaan.
4. Penelitian ini berfokus dalam membahas analisis data pada model rekomendasi buku dengan menggunakan Bahasa pemrograman Python sebagai Machine learning.

C. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

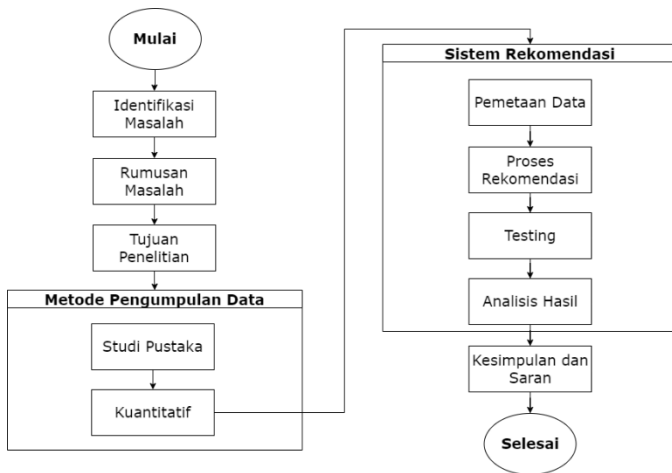
1. Dapat membandingkan dua algoritma perhitungan similarity pada model rekomendasi menggunakan metode item-based collaborative filtering.
2. Memberikan hasil berupa rekomendasi item berdasarkan rating yang diberikan user.

D. Manfaat Penelitian

1. Mengetahui kelebihan dari masing-masing algoritma perhitungan *Cosine Similarity* maupun *Euclidean Distance* pada model rekomendasi buku dengan metode item-based collaborative filtering
2. Memberikan tambahan acuan terhadap pengembangan model rekomendasi khususnya model rekomendasi yang menggunakan metode item-based collaborative filtering.

II. METODOLOGI PENELITIAN

Metodologi merupakan suatu gambaran yang memuat tahapan yang dilaksanakan selama penelitian ini. Gambaran tersebut diperlukan sebagai acuan supaya hasil yang didapatkan memenuhi tujuan dari penelitian ini. Dalam penelitian ini terdapat beberapa tahap dalam pengerjaan untuk digunakan sebagai awal dalam menyelesaikan penelitian. Dalam melakukan penelitian ini, ada beberapa tahapan yang akan dikerjakan meliputi pengumpulan data, pemodelan, perancangan, dan testing seperti yang digambarkan pada gambar berikut:



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1 diatas, penelitian ini memiliki beberapa tahap yang harus dipenuhi sebelum untuk melanjutkan tahapan lainnya. Adapun tahapan dalam penelitian antara lain:

A. Analisa Masalah

Tahap awal pada penelitian adalah analisis masalah. Penulis mengangkat beberapa masalah dan fakta yang terjadi saat ini. Berdasarkan beberapa studi literatur dan artikel, menghasilkan beberapa statement yang mengarah sebuah tren penggunaan platform digital yang mengalami peningkatan selama pembelajaran online. Melihat peluang tersebut, banyak bermunculan *empres emergente edu tech* saat ini. Persaingan yang ketat antar perusahaan terjadi. Perusahaan memerlukan suatu *matrix* yang dapat menjelaskan kinerja suatu aplikasi atau platform yang merka buat. Melalui analisa sentimen akan membantu perusahaan memahami perasaan pengguna terkait produk atau layanan dan apa yang pelanggan inginkan, sehingga membuat perusahaan mampu bersaing di pasaran.

B. Studi Literatur

Tahapan studi literatur dilakukan dengan mempelajari jurnal-jurnal penelitian yang terkait dengan model rekomendasi serta mempelajari literatur yang meliputi konsep perhitungan *Cosine Similarity* dan *Euclidean Distance* dengan menggunakan metode item-based collaborative filtering. Studi literatur dilakukan dengan tujuan untuk mempelajari penelitian terkait yang sudah dilakukan sebelumnya sebagai referensi dan acuan dalam menunjang proses pengembangan penelitian ini.

C. Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini menggunakan teknik pengumpulan data kuantitatif dimana data yang diperoleh didasarkan pada kumpulan data statistik. Kumpulan data yang digunakan dikumpulkan oleh pihak ketiga yang berwenang. Peneliti tidak perlu menyebarkan kuesioner

dan hanya memiliki akses ke dataset dari hasil survei organisasi lain yang terkait dengan masalah yang diteliti. Data dari penelitian ini didapatkan dari Kaggle yang dikumpulkan oleh Cai-Nicolas Ziegler dalam 4 minggu merangkap (Agustus / September 2004) dari komunitas Book-Crossing dengan izin dari Ron Hornbaker, CTO Humankind Systems. Berisi 278.858 pengguna (anonim tetapi dengan informasi demografis) memberikan 1.149.780 ratings (eksplisit / implisit) tentang 271.379 buku.

D. Preprocessing Data

Pada tahap preprocessing dilakukan dengan tujuan untuk mendefinisikan relasi antara satu data dengan data yang lain.

1. User Profile Representation

User Profile Representation didefinisikan sebagai representasi digital eksplisit dari identitas pengguna, sehubungan dengan model operasi, aplikasi perangkat lunak, atau situs web yang dikunjungi. User Profile Representation pada metode item-based collaborative filtering menggunakan user-item rating, dimana proses rekomendasi menggunakan nilai rating dari profil user. Pada proses ini user memberikan nilai rating pada item yang direkomendasikan dalam gauge set (training set) sebagai awal pembentukan dari user profile. Gambaran dari user profile representation dapat dilihat pada Tabel 1 berikut ini.

Tabel 1. User Profile Representation

		Item			
		i_1	i_2	...	i_n
User	u_1	$R_{1,1}$	$R_{1,2}$...	$R_{1,n}$
		$P_{1,1}$	$P_{1,2}$...	$P_{1,n}$
	u_2	$R_{2,1}$	$R_{2,2}$...	$R_{2,n}$
		$P_{2,1}$	$P_{2,2}$...	$P_{2,n}$
	u_m	$R_{m,1}$	$R_{m,2}$...	$R_{m,n}$
		$P_{m,1}$	$P_{m,2}$...	$P_{m,n}$

2. Transpose Matrix

Matriks nilai item ini merupakan hasil dari pivot table yang bertujuan agar data dapat disajikan secara efisien berdasarkan parameter-parameter yang diberikan sesuai dengan kebutuhan data dan memudahkan proses analisis data untuk mengambil suatu keputusan. Sebagai contoh terdapat lima item buku yaitu $i_1, i_2, i_3, \dots, i_n$ dan sepuluh user yaitu $u_1, u_2, u_3, \dots, u_n$ yang disajikan dalam tabel berikut:

Tabel 2. Contoh Matriks User-Item Rating

user/item	i_1	i_2	i_3	i_4	i_5
u_1	5	4		1	4

u_2	4	4	4		3
u_3	5		4	3	5
u_4	1	5	2	1	
u_5	4	4	3	3	3
u_6	5	5	4	1	
u_7	5	4		1	4
u_8	4		3	1	4
u_9	4	2		4	5
u_{10}	5		4	2	5

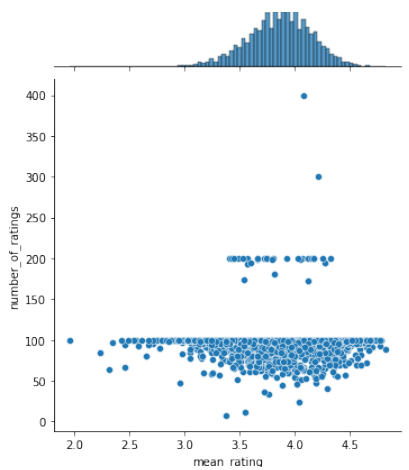
3. Number of Ratings

Pada tahap ini dilakukan filter pada data untuk menggunakan data dengan jumlah rating pada adalah 100. Sehingga setiap item buku memiliki 100 rating dari user. Hal ini dilakukan dengan tujuan untuk mengurangi jarak kelebaran antar item pada data. Untuk detailnya dijelaskan sebagai berikut:

Sebelum dilakukan filter number of rating, dataset memiliki 53424 unique users dan 10000 unique books ditunjukkan pada tabel dan grafik number of rating sebagai berikut.

Tabel 3. Tabel Number of Ratings

title	mean_rating	number_of_ratings
Kindle User's Guide	3.37	8
Diary ng Panget	3.72	11
Fifty Shades Duo: Fifty Shades Darker	3.76	24
Kindle Paperwhite User's Guide	4.04	34
The King's Agent	3.54	36

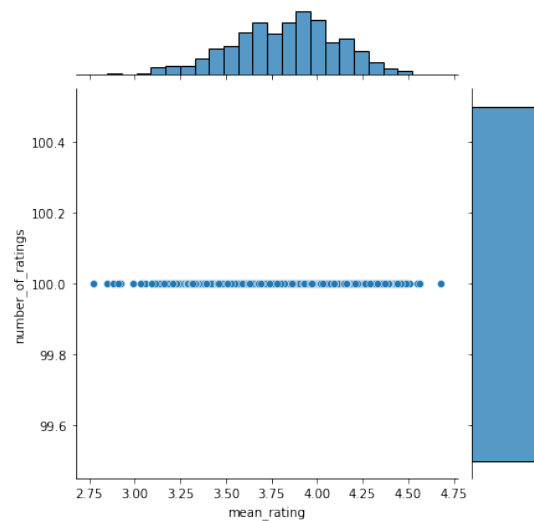


Gambar 2. Number of Ratings

Kemudian berikut kondisi data setelah dilakukan filter dengan number of ratings = 100, dengan dataset memiliki 37233 unique users dan 7400 unique books ditunjukkan pada tabel dan grafik sebagai berikut:

Tabel 4. Tabel Number of Ratings = 100

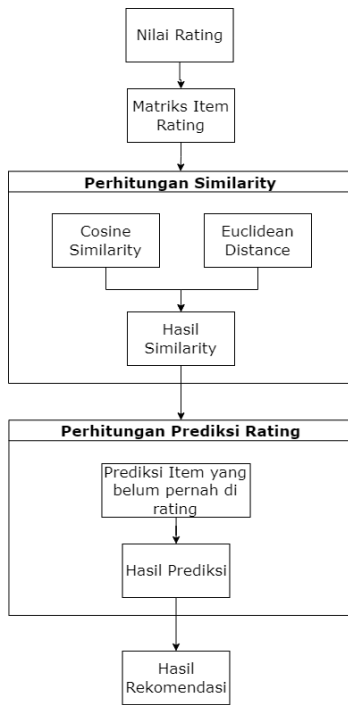
title	mean_rating	number_of_ratings
Angels (Walsh Family, #3)	3.40	100
The Commitments	3.95	100
The Colorado Kid (Hard Case Crime #13)	3.03	100
A Black Man's Tribute. The Color of Water	3.98	100
The Color of Magic Discworld, Rincewind 1	3.80	100



Gambar 3. Number of Ratings = 100

E. Proses Rekomendasi (Processing Stage)

Tahap berikutnya adalah proses dalam melakukan rekomendasi buku, pada proses ini, data rating akan diolah menggunakan metode item-based collaborative filtering –dengan algoritma *Cosine Similarity* dan *Euclidean Distance* hingga memperoleh hasil berupa rekomendasi buku yang diberikan kepada user.



Gambar 2. Proses Rekomendasi

1. Menghitung nilai Similarity Cosine Similarity

Perhitungan Cosine Similarity i_1 dan i_2

$$sim_{jk} = cos_{jk} = \sum_{n=1}^m \frac{r_{jn}}{\sqrt{\sum_n r_{jn}^2}} \frac{r_{kn}}{\sqrt{\sum_n r_{kn}^2}} \quad (1)$$

$$sim_{jk} = \frac{(r_{j1}r_{k1}) + (r_{j2}r_{k2}) + \dots + (r_{jm}r_{km})}{\sqrt{r_{j1}^2 + r_{j2}^2 + \dots + r_{jm}^2} \times \sqrt{r_{k1}^2 + r_{k2}^2 + \dots + r_{km}^2}}$$

$$sim_{1,2} = \frac{(5 * 4) + (4 * 4) + (4 * 5) + (4 * 3) + (5 * 0) + (1 * 5) + (4 * 4) + (5 * 5) + (5 * 4) + (5 * 0)}{\sqrt{5^2 + 4^2 + 4^2 + 4^2 + 5^2 + 1^2 + 4^2 + 5^2 + 5^2 + 5^2} * \sqrt{4^2 + 4^2 + 5^2 + 3^2 + 0^2 + 5^2 + 4^2 + 5^2 + 4^2 + 0^2}}$$

$sim_{1,2} = 0.799092658$

Berikut tabel nilai similarity keseluruhan antar buku yang disajikan dalam tabel 4

Tabel 4. Tabel nilai similarity buku menggunakan Cosine Similarity

Item (buku)	i_1	i_2	i_3	i_4	i_5
i_1	1	0.799092	0.744081	0.725615	0.828951
i_2	0.799092	1	0.543315	0.580343	0.482543
i_3	0.744081	0.543315	1	0.603436	0.680939
i_4	0.725615	0.580343	0.603436	1	0.657873
i_5	0.828951	0.482543	0.680939	0.657873	1

2. Menghitung nilai Similarity Euclidean Distance

a. Perhitungan jarak item Euclidean Distance i_1 dan i_2

$$d_{(i,j)} = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{in} - x_{jn}|^2} \quad (2)$$

$$d_{(i_1,i_2)} = \sqrt{(5 - 4)^2 + (4 - 4)^2 + (4 - 5)^2 + (4 - 3)^2 + (5 - 0)^2 + (1 - 5)^2 + (4 - 4)^2 + (5 - 5)^2 + (5 - 4)^2 + (5 - 0)^2}$$

$$d_{(i_1,i_2)} = \sqrt{70}$$

$$d_{(i_1,i_2)} = 8.366600265$$

Tabel 5. Tabel nilai jarak menggunakan Euclidean Distance

Dist	i_1	i_2	i_3	i_4	i_5
i_1	0	8.366600	9.327379	10.67707	7.9372539
i_2	8.366600	0	10.44030	10.09950	11.269427
i_3	9.327379	10.44030	0	7	7.3484692
i_4	10.67707	10.09950	7	0	7.2801098
i_5	7.937253	11.26942	7.348469	7.280109	0

b. Min-max Normalization

Setelah itu nilai dari jarak antar item buku dinormalisasi dengan skala rentang nilai 0-1 menggunakan metode min-max normalization.

$$d' = \frac{d - \min(p)}{\max(p) - \min(p)} \quad (3)$$

$$d' = \frac{8.366600265 - 0}{11.26942767 - 0}$$

$d' = 0.74241572$

Berikut tabel nilai jarak setelah dinormalisasi menggunakan min-max Normalization

Tabel 6. Tabel nilai jarak setelah dinormalisasi

Min-max Dist	i_1	i_2	i_3	i_4	i_5
i_1	0	0.7424157	0.8276710	0.9474374	0.7043173
i_2	0.7424157	0	0.9264273	0.8961861	1
i_3	0.8276710	0.9264273	0	0.6211495	0.6520712
i_4	1	0.8961861	0.6211495	0	0.6460052
i_5	0.7043173	1	0.6520712	0.6460052	0

c. Nilai Similarity Euclidean Distance

Setelah selesai dinormalisasi maka tahapan selanjutnya menghitung nilai similarity antar item menggunakan persamaan seperti di bawah ini.

$$sim(x, y) = \frac{1}{1 - d'(x, y)} \quad (4)$$

Dengan $sim(x,y)$ adalah nilai kesamaan antar film dan $d'(x,y)$ adalah hasil perhitungan Euclidean Distance. Berikut perhitungan similaritas antar item buku untuk contoh kasus di atas.

$$sim(i_1, i_2) = \frac{1}{1 - 0.742415721}$$

$$sim(i_1, i_2) = \mathbf{0.57391585}$$

Berikut tabel hasil perhitungan similarity antar item buku.

Tabel 7. Tabel nilai similarity item buku menggunakan Euclidean Distance

Similarity	i_1	i_2	i_3	i_4	i_5
i_1	1	0.5739158	0.5471444	0.5134953	0.5867451
i_2	0.5739158	1	0.5190956	0.5273743	0.5
i_3	0.5471444	0.5190956	1	0.6168462	0.6053007
i_4	0.5134953	0.5273743	0.6168462	1	0.6229362
i_5	0.5867451	0.5	0.6053007	0.6075314	1

F. Prediksi Rating

Menghitung Prediksi Rating dengan $S_{i,j}$ adalah nilai similarity item i dan item j , $R_{u,j}$ adalah nilai rating user u pada item I dan $P_{(u,i)}$ adalah nilai prediksi rating user u pada item. Berikut perhitungan prediksi rating dari contoh kasus di atas menggunakan *Cosine Similarity*.

$$P_{(u,i)} = \frac{\sum_{j \in (R_{u,j} * S_{i,j})}}{\sum_{j \in I} |S_{i,j}|} \quad (5)$$

- Prediksi u_1, i_3

$$P_{u_1, i_3} = \frac{(5 * 0.744081225) + (4 * 0.543315098) + (1 * 0.603436711) + (4 * 0.680939922)}{(0.744081225 + 0.543315098 + 0.603436711 + 0.680939922)}$$

$$P_{u_1, i_3} = \mathbf{3.58541095}$$

$$P_{u_2, i_4} = \mathbf{3.743745717}$$

$$P_{u_3, i_3} = \mathbf{4.422026293}$$

$$P_{u_3, i_4} = \mathbf{4.444380928}$$

$$P_{u_3, i_5} = \mathbf{4.367933989}$$

$$P_{u_4, i_4} = \mathbf{3.074382302}$$

$$P_{u_4, i_5} = \mathbf{3.086869184}$$

$$P_{u_5, i_2} = \mathbf{4.291562158}$$

$$P_{u_6, i_5} = \mathbf{1.985211312}$$

$$P_{u_8, i_5} = \mathbf{3.750169952}$$

$$P_{u_9, i_3} = \mathbf{3.58541095}$$

$$P_{u_{10}, i_2} = \mathbf{4.437870685}$$

$$P_{u_{10}, i_4} = \mathbf{4.365195103}$$

Berikut perhitungan prediksi rating dari contoh kasus di atas menggunakan *Euclidean Distance*.

- Prediksi u_1, i_3

$$P_{u_1, i_3} = \frac{(5 * 0.54714441) + (4 * 0.519095608) + (1 * 0.616846235) + (4 * 0.605300788)}{(0.54714441 + 0.519095608 + 0.616846235 + 0.605300788)}$$

$$P_{u_1, i_3} = \mathbf{3.430431011}$$

$$P_{u_2, i_4} = \mathbf{3.726860486}$$

$$P_{u_3, i_3} = \mathbf{4.486846863}$$

$$P_{u_3, i_4} = \mathbf{4.506667061}$$

$$P_{u_3, i_5} = \mathbf{4.460089461}$$

$$P_{u_4, i_4} = \mathbf{2.958586897}$$

$$P_{u_4, i_5} = \mathbf{2.987731465}$$

$$P_{u_5, i_2} = \mathbf{4.257755667}$$

$$P_{u_6, i_5} = \mathbf{2.13294763}$$

$$P_{u_8, i_5} = \mathbf{3.680006766}$$

$$P_{u_9, i_3} = \mathbf{3.430431011}$$

$$P_{u_{10}, i_2} = \mathbf{4.360271012}$$

$$P_{u_{10}, i_4} = \mathbf{4.292877327}$$

G. Rekomendasi

Setelah hasil dari prediksi dari rating kosong pada masing-masing rumus perhitungan *Cosine Similarity* dan *Euclidean Distance* didapatkan, selanjutnya data dari hasil prediksi item akan diurutkan dari nilai prediksi tertinggi ke yang terendah terhadap user. Dan dari hasil tersebut adalah rekomendasi item kepada user dari item yang belum di rating oleh user tersebut berdasarkan item-item lain yang telah di rating. Berikut hasil rekomendasi yang diberikan dengan rumus perhitungan *Cosine Similarity*.

- $u_1 = \{i_3\}$
- $u_2 = \{i_4\}$
- $u_3 = \{i_3, i_4, i_5\}$
- $u_4 = \{i_5, i_4\}$
- $u_5 = \{i_2\}$
- $u_6 = \{i_5\}$
- $u_8 = \{i_5\}$
- $u_9 = \{i_3\}$
- $u_{10} = \{i_2, i_4\}$

Dan berikut hasil rekomendasi yang diberikan dengan rumus perhitungan *Euclidean Distance*.

- $u_1 = \{i_3\}$
- $u_2 = \{i_4\}$

- $u_3 = \{i_4, i_3, i_5\}$
- $u_4 = \{i_5, i_4\}$
- $u_5 = \{i_2\}$
- $u_6 = \{i_5\}$
- $u_8 = \{i_5\}$
- $u_9 = \{i_3\}$
- $u_{10} = \{i_2, i_4\}$

H. Post-processing

Tujuan dari pengujian kualitas rekomendasi adalah untuk mengetahui algoritma perhitungan similarity mana yang menghasilkan kualitas rekomendasi yang lebih baik yang dapat dilihat dari keakuratan nilai prediksi yang dihasilkan oleh masing-masing algoritma. Pada pengujian ini juga digunakan metode k-fold cross validation dengan nilai k=10 sehingga semua data rating yang didapat dibagi menjadi 10 dataset. Satu dataset digunakan untuk testing dan dataset-dataset yang lain digunakan untuk training. Lalu setelah itu nilai MAE pada tiap dataset akan dirata-rata sehingga nilai rata-rata MAE tersebut menjadi perwakilan nilai akurasi dari perhitungan similarity yang digunakan pada penelitian ini. Penggunaan metode k-10 cross validation ini dikarenakan menurut (Kohavi, 1995) jumlah k-10 merupakan jumlah yang paling direkomendasikan untuk mengukur tingkat validasi. Kemudian pengujian kualitas rekomendasi dapat dilakukan dengan metric evaluation method, salah satunya adalah Mean Absolute Error (MAE) seperti yang digunakan pada penelitian. Kualitas rekomendasi diukur dengan parameter akurasi dari prediksi nilai rating yang dihasilkan oleh model.

I. Analisis Hasil

Dari hasil nilai MAE yang diperoleh indikator pengujian dapat dilihat dari semakin kecil nilai MAE maka semakin baik kualitas rekomendasi yang dihasilkan oleh model. Kemudian dibuat analisis berdasarkan hasil pengujian yang diperoleh dari dengan perhitungan similarity *Cosine Similarity* dan *Euclidean Distance* pada model rekomendasi Item-based Collaborative Filtering.

III. HASIL DAN PEMBAHASAN

A. Hasil Rekomendasi *Cosine Similarity*

Tahap berikutnya adalah proses dalam melakukan rekomendasi buku, pada proses ini, data rating akan diolah menggunakan metode *Item-based Collaborative Filtering* - dengan algoritma *Cosine Similarity*. Rekomendasi yang diberikan kepada tiap user merupakan 10 item buku dengan rating prediksi tertinggi. Dalam proses untuk melakukan rekomendasi, pertama dilakukan penghitungan dengan menggunakan rumus *Cosine Similarity*. Kemudian perhitungan prediksi dilakukan dengan menggunakan metode *Weighted Sum*. Berikut adalah hasil berupa daftar buku yang direkomendasikan kepada user menggunakan perhitungan *Cosine Similarity* untuk user_id=588:

Tabel 4. 1 Tabel Hasil Rekomendasi Buku menggunakan *Cosine Similarity*

No.	Buku	Prediksi Rating
1	Snow Falling on Cedars	4.29967947636
2	The Shipping News	4.295035843446
3	The Historian	4.293400439701
4	Cold Mountain	4.220209991216
5	The Alienist (Dr. Laszlo Kreizler, #1)	4.194599197968
6	A Fine Balance	4.187352128685
7	The Very Hungry Caterpillar Board Book	4.158474064865
8	Prey	4.121199030220
9	The Waste Lands (The Dark Tower, #3)	4.119374305214
10	Never Let Me Go	4.119374305214922

B. Hasil Rekomendasi *Euclidean Distance*

Rekomendasi yang diberikan kepada tiap user merupakan 10 item buku dengan rating prediksi tertinggi. Perhitungan prediksi dilakukan dengan menggunakan metode *Weighted Sum*. Berikut adalah hasil rekomendasi menggunakan perhitungan *Euclidean Distance* untuk user_id=588:

Tabel 4. 2 Tabel Hasil Rekomendasi Buku menggunakan *Euclidean Distance*

No.	Buku	Prediksi Rating
1	Snow Falling on Cedars	4.002568564526
2	The Shipping News	4.099554781083
3	The Historian	2.6955440135768
4	Cold Mountain	3.9020190664648
5	The Alienist (Dr. Laszlo Kreizler, #1)	4.101715637212
6	A Fine Balance	3.802564473910
7	The Very Hungry Caterpillar Board Book	2.600224938641
8	Prey	2.391060981743
9	The Waste Lands (The Dark Tower, #3)	2.395876742838
10	Never Let Me Go	3.802065141812

C. Post-Processing

1. Pengujian

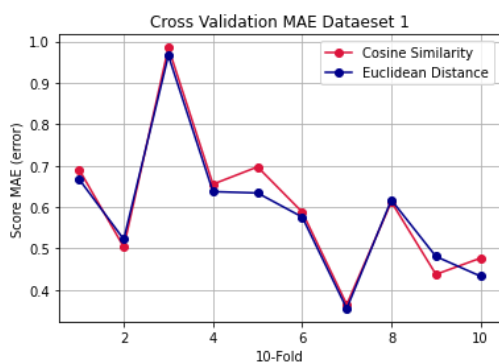
Dari data sampel yang sudah didapat, lalu dibagi menjadi 5 dataset dari total 500 data rating. Pada pengujian ini akan menggunakan metode k-10 cross validation sehingga ada 10 kali pengujian. Satu dataset akan digunakan untuk testing dan dataset-dataset yang lain akan digunakan untuk training. Untuk setiap dataset yang akan dihitung prediksi dengan metode item-based collaborative filtering yang menggunakan perhitungan *Cosine Similarity* dan perhitungan *Euclidean Distance*. Kemudian dilakukan pengujian menggunakan metode Mean Absolute Error (MAE). Dengan $\hat{e}_{u,i}$ adalah nilai prediksi rating pengguna u pada item i, dan $e_{u,i}$ adalah nilai rating yang sebenarnya pengguna u pada item i maka persamaan MAE dirumuskan sebagai berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n | \hat{e}_{u,i} - e_{u,i} | \quad (6)$$

Di mana n adalah jumlah item yang telah di-rating oleh pengguna. Semakin kecil nilai MAE, semakin akurat model dalam memberikan rekomendasi. sehingga didapatkan hasil seperti di Tabel dan Gambar berikut.

Tabel 8. Hasil Cross Validation MAE dataset 1

Fold	Nilai MAE <i>Cosine Similarity</i>	Nilai MAE <i>Euclidean Distance</i>
1	0.6902725	0.6669523
2	0.5033523	0.5221449
3	0.9850240	0.9657865
4	0.6551890	0.6372893
5	0.6969747	0.6339366
6	0.5881539	0.5759999
7	0.3648611	0.3550544
8	0.6125535	0.6171412
9	0.4375291	0.4807209
10	0.4767126	0.4331003
Rata-rata	0.601062312	0.588812655



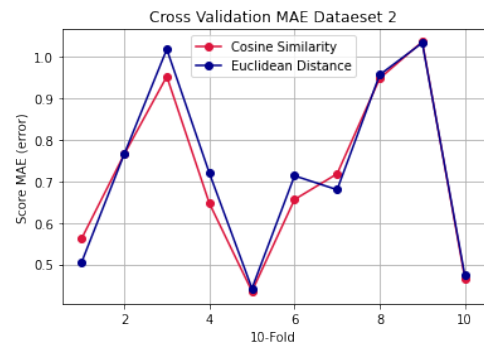
Gambar 4. 1 Grafik hasil Pengujian MAE

Pengujian cross validation pada dataset pertama menunjukkan nilai error MAE yang relatif dekat, dimana pada dataset pertama memiliki nilai sparsity sebesar 25%. Jika dilihat kembali pada grafik hasil

pengujian menunjukkan nilai error MAE dengan perhitungan similarity *Euclidean Distance* lebih rendah daripada *Cosine Similarity*.

Tabel 9. Hasil Cross Validation MAE dataset 2

Fold	Nilai MAE <i>Cosine Similarity</i>	Nilai MAE <i>Euclidean Distance</i>
1	0.5622869	0.5051733
2	0.7652499	0.7659415
3	0.9531663	1.0187665
4	0.6470942	0.7199925
5	0.4349210	0.4419156
6	0.6563665	0.7135789
7	0.7183292	0.6799162
8	0.9486044	0.9572388
9	1.0369683	1.0342041
10	0.4663622	0.4727494
Rata-rata	0.71893493	0.7309477



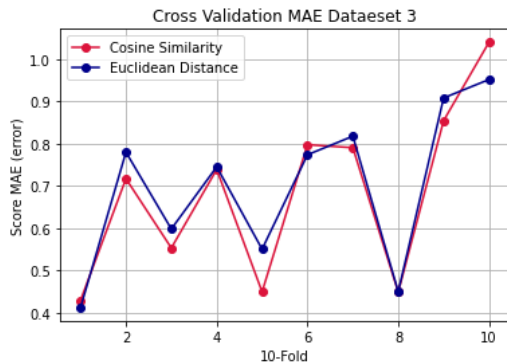
Gambar 4. Grafik Cross Validation MAE dataset 2

Pengujian cross validation pada dataset ke-2 menunjukkan nilai error MAE yang relatif tinggi, dimana pada fold ke-3 dan ke-8 memiliki nilai error mencapai angka 1. Pada dataset ke-2 memiliki nilai sparsity sebesar 28%. Jika dilihat kembali pada grafik hasil pengujian menunjukkan nilai error MAE dengan perhitungan similarity *Cosine Similarity* lebih rendah daripada *Euclidean Distance*.

Tabel 10. Hasil Cross Validation MAE dataset 3

Fold	Nilai MAE <i>Cosine Similarity</i>	Nilai MAE <i>Euclidean Distance</i>
1	0.4293998	0.4124079
2	0.7171221	0.7798241
3	0.5531678	0.5981100
4	0.738932	0.7451431
5	0.4495295	0.5503837
6	0.7974916	0.7736982
7	0.7974916	0.8175962
8	0.4489017	0.4499682
9	0.4489017	0.9081008
10	1.0398348	0.9516991

Rata-rata	0.68192135	0.698693175
-----------	-------------------	--------------------

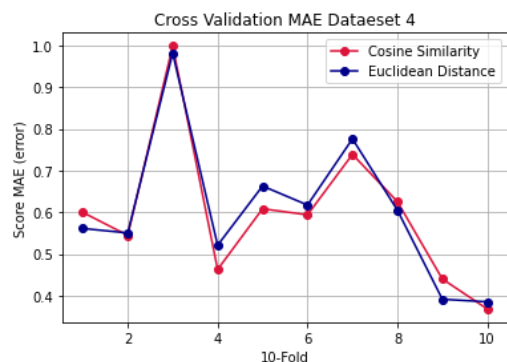


Gambar 5. Grafik Cross Validation MAE dataset 3

Pengujian cross validation pada dataset ke-3 menunjukkan nilai rata-rata error MAE pada masing-masing perhitungan similarity memiliki nilai yang mirip. Dimana pada dataset ke-3 memiliki nilai sparsity sebesar 28,9%. Jika dilihat kembali pada grafik hasil pengujian menunjukkan nilai error MAE dengan perhitungan similarity *Cosine Similarity* lebih rendah daripada *Euclidean Distance*.

Tabel 11. Hasil Cross Validation MAE dataset 4

Fold	Nilai MAE <i>Cosine Similarity</i>	Nilai MAE <i>Euclidean Distance</i>
1	0.59971244	0.56101227
2	0.54448673	0.55037081
3	0.99928905	0.98211171
4	0.46261775	0.52079461
5	0.60834108	0.66290715
6	0.60834108	0.6172796
7	0.59374467	0.77617922
8	0.73885693	0.60493433
9	0.43983015	0.39113157
10	0.36769574	0.38491965
Rata-rata	0.5980046423	0.6051640924

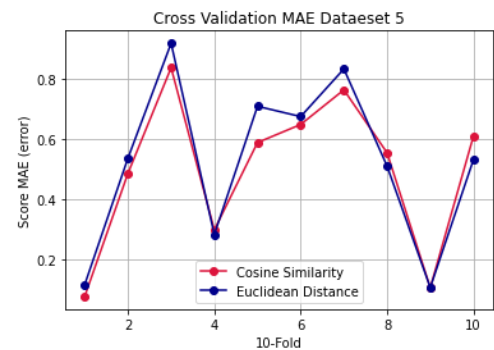


Gambar 6. Grafik Cross Validation MAE dataset 4

Pengujian cross validation pada dataset ke-4 menunjukkan nilai rata-rata error MAE pada masing-masing perhitungan similarity memiliki nilai yang mirip. Dimana pada dataset ke-4 memiliki nilai sparsity sebesar 28%. Jika dilihat kembali pada grafik hasil pengujian menunjukkan nilai error MAE dengan perhitungan similarity *Cosine Similarity* lebih rendah daripada *Euclidean Distance*.

Tabel 12. Hasil Cross validation Mean Absolute Error dataset 5

Fold	Nilai MAE <i>Cosine Similarity</i>	Nilai MAE <i>Euclidean Distance</i>
1	0.0759006	0.1152765
2	0.4857631	0.5365132
3	0.8380788	0.9194797
4	0.2978004	0.2822384
5	0.5885796	0.7095242
6	0.6489204	0.6752254
7	0.7643112	0.8339450
8	0.5545869	0.5101695
9	0.1047568	0.1056131
10	0.6100850	0.5317684
Rata-rata	0.4968783248	0.5219753823



Gambar 7. Grafik Cross Validation MAE dataset 5

Pengujian cross validation pada dataset ke-5 menunjukkan nilai error MAE yang relatif rendah, dimana pada fold ke-1 dan ke-0 memiliki nilai error mendekati angka 0. Pada dataset ke-5 memiliki nilai sparsity sebesar 23%. Jika dilihat kembali pada grafik hasil pengujian menunjukkan nilai error MAE dengan perhitungan similarity *Cosine Similarity* lebih rendah daripada *Euclidean Distance*.

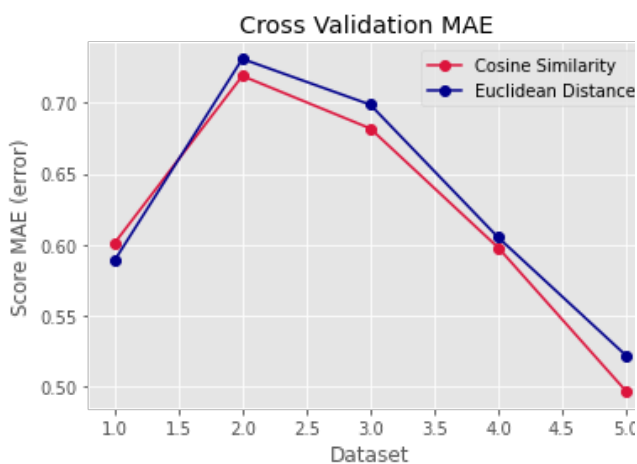
2. Analisis Hasil Pengujian

Pada tahap sebelumnya telah dilakukan pengujian dan perhitungan menggunakan MAE pada rumus perhitungan *Cosine Similarity* dan *Euclidean Distance* dengan metode item-based collaborative filtering. Data rating yang digunakan adalah 500 data rating yang dibagi menjadi 5 dataset, dimana satu dataset

memiliki 100 data rating. Pembagian pada dataset yang akan diuji dilakukan untuk melihat perbandingan bentuk grafik yang dihasilkan sehingga terlihat perbedaan nilai MAE pada setiap dataset dan hasil dari rata-rata nilai MAE, selain itu untuk menunjukkan nilai MAE setiap fold pada setiap dataset. Kemudian pada setiap dataset digunakan sebagai hasil perbandingan dari hasil pengujian pada masing-masing rumus perhitungan yang diuji. Tabel 13. dan Gambar 8. ini menunjukkan perbandingan nilai MAE yang dihasilkan oleh rumus perhitungan *Cosine Similarity* dan *Euclidean Distance* dengan metode item-based collaborative filtering.

Tabel 13. Hasil testing Mean Absolute Error

dataset	Nilai MAE <i>Cosine Similarity</i>	Nilai MAE <i>Euclidean Distance</i>
1	0.601062312233405	0.588812655170038
2	0.718934930573236	0.730947724223521
3	0.681921351137601	0.698693175641707
4	0.598004642361587	0.605164092432937
5	0.496878324833384	0.521975382323462
Rata-rata	0.647351928695726	0.676872059975026



Gambar 8. Grafik hasil Pengujian MAE

Setelah dilakukan pengujian menggunakan rumus MAE, menunjukkan bahwa dengan menggunakan rumus kesamaan item *Cosine Similarity* memiliki hasil kualitas rekomendasi yang lebih baik dibandingkan dengan rumus *Euclidean Distance* pada metode *Item-based Collaborative Filtering*. Hal tersebut dapat diketahui dengan melihat nilai MAE yang dihasilkan oleh masing-masing rumus perhitungan. Nilai MAE yang dihasilkan oleh rumus perhitungan *Cosine Similarity* lebih kecil daripada nilai MAE yang dihasilkan oleh rumus perhitungan *Euclidean Distance* dengan perbedaan angkanya cukup dekat. Semakin kecil nilai MAE yang

dihasilkan maka semakin baik pula kualitas rekomendasi dan akurasi prediksi yang dihasilkan oleh model. Dari tabel dan grafik di atas, terlihat bahwa model rekomendasi item-based Collaborative Filtering yang menggunakan metode *Cosine Similarity* dan *Euclidean Distance* mengalami titik tertinggi nilai MAE saat pengujian dataset ke-2 dan titik terendah nilai MAE pada saat pengujian dataset ke-5.

Jika dilihat kembali pada grafik hasil pengujian, nilai MAE dari model rekomendasi dengan rumus perhitungan *Cosine Similarity* dan *Euclidean Distance* membentuk grafik yang mirip karena memiliki selisih nilai MAE yang dekat pada setiap dataset. Namun dikarenakan hasil nilai rata-rata MAE pada perhitungan similarity *Euclidean Distance* dan *Cosine Similarity* memiliki nilai yang mirip maka bisa disimpulkan bahwa pada penelitian ini, model rekomendasi menggunakan perhitungan *Cosine Similarity* sedikit lebih unggul dan memiliki tingkat akurasi yang lebih baik dibandingkan dengan perhitungan *Euclidean Distance*. Semakin baik akurasi prediksi yang dihasilkan maka semakin baik pula kualitas rekomendasinya.

IV. KESIMPULAN

1. Dengan perbedaan dari hasil tingkat akurasi yang tidak terlalu jauh diantara kedua rumus perhitungan *similarity*, rumus perhitungan *Cosine Similarity* lebih berkualitas dalam menghasilkan rekomendasi. Dari hasil pengujian tersebut dapat disimpulkan bahwa rumus perhitungan yang lebih efektif untuk model rekomendasi buku pada penelitian ini adalah *Cosine Similarity*.
2. Perbandingan dari hasil pengujian *Mean Absolute Error* (MAE) dengan *10-Fold Cross Validation* dapat disimpulkan bahwa rumus perhitungan *Cosine Similarity* mempunyai tingkat akurasi yang cukup baik yaitu 0.647352 dibandingkan dengan rumus perhitungan *Euclidean Distance* memiliki nilai akurasi yaitu 0.676872 dengan skala MAE 0–1. Semakin kecil nilai MAE semakin tinggi akurasi prediksi yang dihasilkan. Hal ini menunjukkan bahwa rumus perhitungan *Cosine Similarity* dapat menghasilkan rekomendasi yang lebih baik dari *Euclidean Distance*. Hal ini disebabkan karena pada perhitungan prediksi dari metode *Euclidean Distance* memiliki nilai yang lebih jauh dengan rating asli maupun saat data yang masih sedikit maupun saat data bertambah banyak yang mengakibatkan nilai MAE yang dihasilkan bertambah besar, sebaliknya pada hasil nilai prediksi yang dihasilkan oleh rumus perhitungan *Cosine Similarity* memiliki nilai prediksi yang tidak terlalu jauh dengan rating asli.

V. SARAN

Berdasarkan kesimpulan di atas dapat diberikan saran untuk penelitian selanjutnya yaitu:

1. Perlunya dilakukan perbandingan pengujian lebih lanjut menggunakan metode perhitungan similaritas lain seperti Pearson Correlation untuk dapat mengetahui metode perhitungan mana yang lebih baik untuk diimplementasikan pada model rekomendasi buku Item-based Collaborative Filtering.
2. Pada pengujian kali ini menggunakan metode Item-based Collaborative Filtering dengan menghitung kesamaan antar item sehingga penulis berharap ada penelitian lebih lanjut terkait kasus model rekomendasi Collaborative Filtering dengan menghitung kesamaan antar user.

UCAPAN TERIMA KASIH

Ucapan terimakasih penulis ditujukan kepada Tuhan Yang Maha Esa, atas karunia-Nya sehingga penulis dapat menyelesaikan laporan penelitian dengan baik dan lancar. Tidak lupa banyak terimakasih ini diucapkan dan ditunjukkan kepada kedua orang tua, dan kepada Ibu Yuni Yamasari sebagai Dosen Pembimbing skripsi, serta teman-teman dan semua pihak yang sudah terlibat dalam selesainya penelitian ini.

REFERENSI

- Alharthi, H., Inkpen, D., Szpakowicz, S. (2017). A Survey of Book Recommender Systems. *Journal of Intelligent Information Systems*, 139–160.
- E. Uko Okon, B. O. Eke, P. O. Asagba. (2018). An Improved Online Book Recommender System using Collaborative Filtering Algorithm. *International Journal of Computer Applications*, 0975 – 8887.
- Gupta, M., Thakkar, A., Gupta, V., & Rathore, D. P. S. . (2020). Movie recommender system using collaborative filtering. *International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 415-420.
- H. Koochi, K. Kiani. (2017). A New Method to find Neighbor Users that Improves the Performance. *Expert Systems With Applications*, vol. 83, 30–39.
- Halim, Y. A. (2013). Digital Cakery Dengan Algoritma Collaborative Filtering. *Jurnal SIFO Mikroskil*, Vol 14, No 1, 79-88.
- Hikmatyar, M., Ruuhwan, Zery H. (2019). Sistem Rekomendasi Buku di Perpustakaan Berbasis User-Based Collaborative Filtering. *Jurnal ICT : Information Communication & Technology*, Vol.18, No.2, 71-77.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*, Vol. 14, No. 2, pp. 1137-1145.
- Laja Jaja, Y. V., Bambang, S., Leopoldus, R. S. (2020). Penerapan Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Data MovieLens. *Jurnal Matematika dan Aplikasi*, Vol. 9, No. 2, 78-83.
- Mustofa, A. R. (2018). Sistem Rekomendasi Wisata Di Kediri Menggunakan Metode Item Based Collaborative Filtering. Universitas Islam Negeri Maulana Malik Ibrahim.
- Sari, S., & Hendra, D. T. . (2015). Aplikasi Rekomendasi Film menggunakan Pendekatan Collaborative Filtering dan *Euclidean Distance* sebagai ukuran kemiripan rating. *Seminar Nasional Teknologi Informasi dan Komunikasi Terapan (SEMANTIK)*, 135-140.
- Sarwar et al. (2001). Item-Based Collaborative Filtering Recommendation Algorithms. *Proceedings of the 10th*, 285-295.
- Wijaya, A. E., Deni A. (2018). Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering dan Content-Based Filtering. *Jurnal Computech & Bisnis*, Vol 12, No 1, 11-27.