

**ANALISIS PERBANDINGAN COSINE DAN PEARSON UNTUK
MENGOPTIMALKAN TEKNIK COLLABORATIVE FILTERING PADA
SISTEM REKOMENDASI DI DOMAIN BUKU**

SKRIPSI

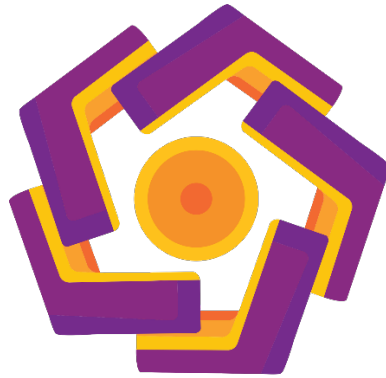


**disusun oleh
Adnan Aziz Dhia Azzindani
17.11.1016**

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**ANALISIS PERBANDINGAN COSINE DAN PEARSON UNTUK
MENGOPTIMALKAN TEKNIK COLLABORATIVE FILTERING PADA
SISTEM REKOMENDASI DI DOMAIN BUKU**

SKRIPSI



**disusun oleh
Adnan Aziz Dhia Azzindani
17.11.1016**

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**ANALISIS PERBANDINGAN COSINE DAN PEARSON UNTUK
MENGOPTIMALKAN TEKNIK COLLABORATIVE FILTERING PADA
SISTEM REKOMENDASI DI DOMAIN BUKU**

yang dipersiapkan dan disusun oleh

Adnan Aziz Dhia Azzindani

17.11.1016

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 8 October 2021

Dosen Pembimbing,

Arif Dwi Laksito, M.Kom

NIK. 190302150

PENGESAHAN

SKRIPSI

**ANALISIS PERBANDINGAN COSINE DAN PEARSON UNTUK
MENGOPTIMALKAN TEKNIK COLLABORATIVE FILTERING
PADA SISTEM REKOMENDASI DI DOMAIN BUKU**

yang dipersiapkan dan disusun oleh

Adnan Aziz Dhia Azzindani

17.11.1016

telah dipertahankan di depan Dewan Penguji

pada tanggal 7 October 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Mardhiya Hayaty, S.T.,M.Kom.

NIK : 190302108

Hartatik, S.T., M.Cs.

NIK : 190302232

Arif Dwi Laksito, M.Kom

NIK. 190302150

Skripsi ini telah diterima sebagai salah satu persyaratan

untuk memperoleh gelar Sarjana Komputer

Tanggal 04 November 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom.

NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 12 Desember 2021



Adnan Aziz DhiaAzzindani

NIM. 17.11.1016

MOTTO

“Jangan pernah berhenti berjuang sampai titik darah penghabisan sebelum Tuhan bilang sudah waktunya pulang.”

(Adnan Aziz D)

“Dan barangsiapa berusaha, maka sesungguhnya usahanya itu untuk dirinya sendiri”
(QS 29:6)



PERSEMBAHAN

Puji syukur kita panjatkan kehadiran Allah SWT atas berkah dan karunia-Nya skripsi ini dapat terselesaikan dengan baik dan lancar. Dengan ini saya persembahkan skripsi ini untuk semua pihak yang terlibat langsung maupun tidak langsung, yaitu kepada :

1. Kedua orang tua saya, kakak, adik, sodara maupun ayah sambung saya dan keluarga besar yang selalu mendoakan, merestui dan memberikan motivasi tiada henti.
2. Pacar saya Sonia Purnama Putri yang sudah memberikan semangat dan support tanpa henti.
3. Bapak Arif Dwi Laksito, M.Kom yang telah mengenalkan saya kepada dunia pemrograman yang lebih luas dan telah bersabar membimbing saya dari awal hingga akhir.
4. Dosen-dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu semasa kuliah.
5. Teman-teman kost yang selalu mengingatkan kapan selesai skripsinya.
6. Terimakasih yang sebesar-besarnya untuk diri saya sendiri karena telah berjuang sampai dititik ini.

KATA PENGHANTAR

Puji syukur penulis ucapkan kepada Allah SWT, yang mana telah memberikan kesehatan dan karunia-Nya kepada penulis serta kekuatan untuk menyelesaikan skripsi yang berjudul “ **ANALISIS PERBANDINGAN COSINE DAN PEARSON UNTUK MENGOPTIMALKAN TEKNIK COLLABORATIVE FILTERING PADA SISTEM REKOMENDASI DI DOMAIN BUKU**”. Tidak lupa penulis mengucapkan shalawat dan salam kepada junjungan Nabi Besar Muhammad SAW. Penyelesaian tulisan ini terlepas bantuan dari berbagai pihak yang terkait secara langsung maupun tidak langsung, terutama dan teristimewa dipersembahkan kepada kedua orang tua tercinta yang senantiasa memberikan rasa sayang, didikan, serta doa yang selalu di panjatkan pada Allah kepada penulis.

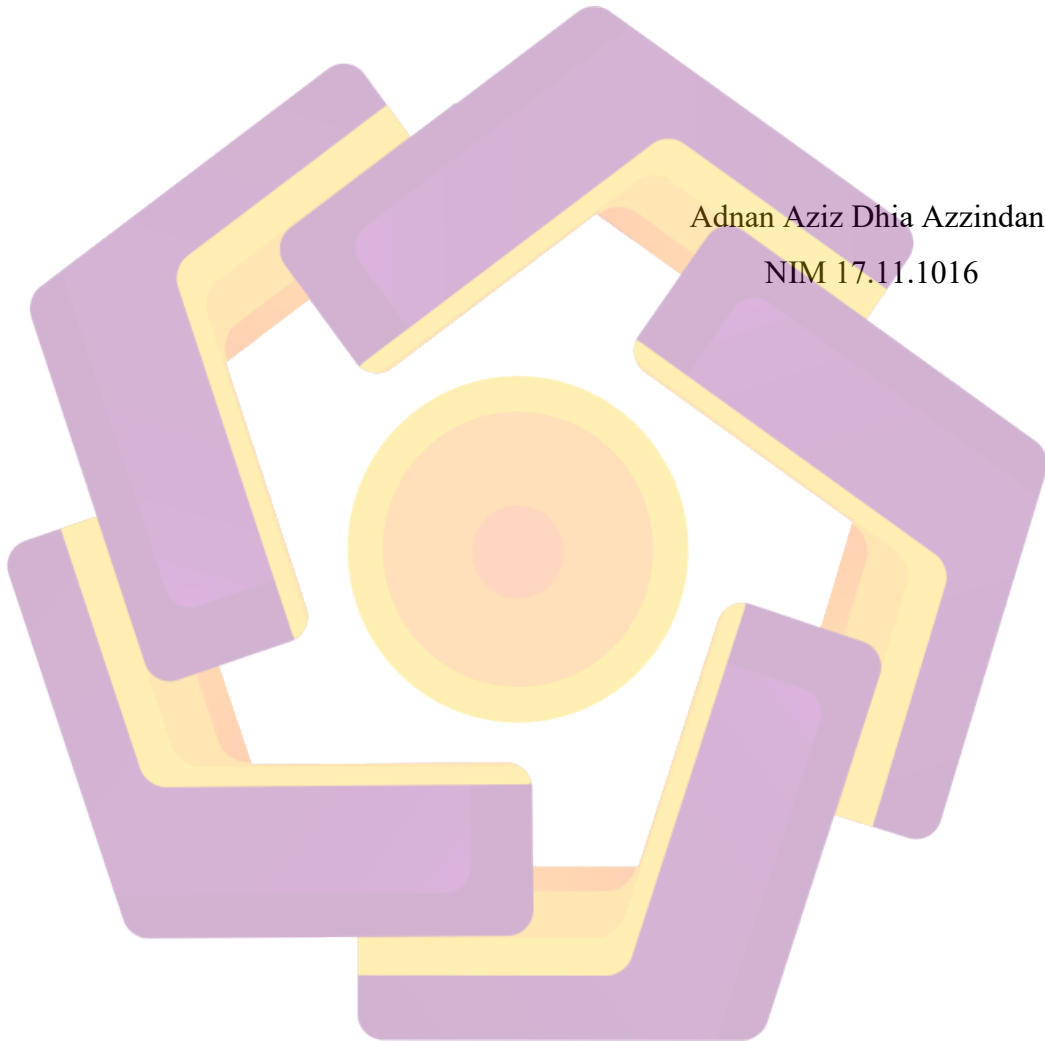
Skripsi ini dapat terselesaikan dengan bantuan berbagai pihak, maka dari itu penulis menyatakan rasa hormat dan terimakasih kepada:

1. Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradnya D, M.Kom selaku ketua Program Studi Informatika Universitas AMIKOM Yogyakarta.
4. Arif Dwi Laksito, M.Kom selaku pembimbing yang senantiasa memberikan masukan serta nasihat dalam penulisan skripsi ini.
5. Ibu Mardhiya Hayaty, S.T., M.Kom, dan Ibu Hartatik, S.T.,M.Cs. Selaku dosen penguji, terima kasih atas saran dan kritiknya sehingga penelitian ini menjadi lebih baik lagi.

Penulis menyadari masih ada kekurangan dan kelemahan dalam pembuatan skripsi ini. Maka penulis mengharapkan adanya kritik dan saran dari segala pihak agar menambah kesempurnaan dalam skripsi ini.

Yogyakarta, 04 November 2021

Adnan Aziz Dhia Azzindani
NIM 17.11.1016

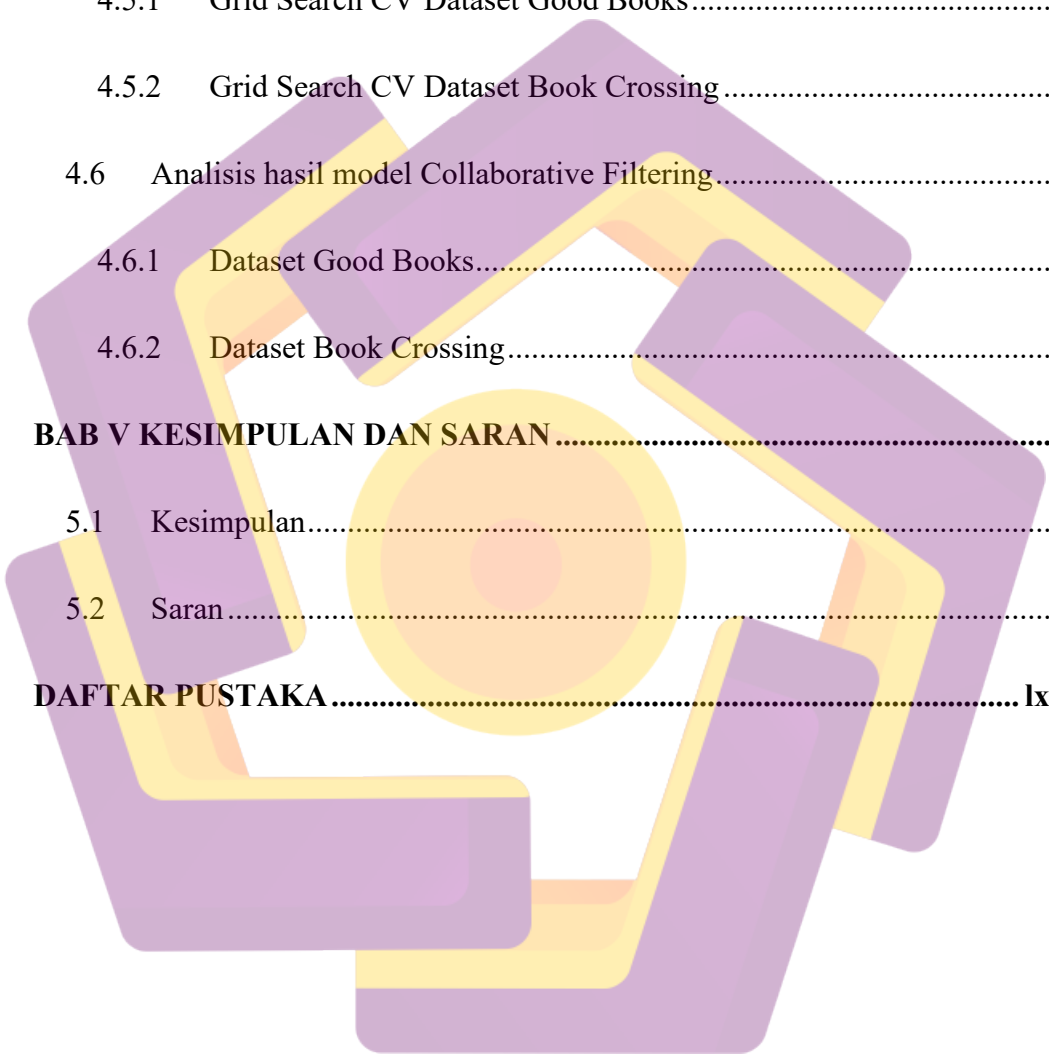


DAFTAR ISI

PERSETUJUAN	iii
PENGESAHAN	iv
PERNYATAAN	v
MOTTO	vi
PERSEMBAHAN	vii
KATA PENGHANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xvi
INTISARI	xvii
ABSTRACT	xviii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan dan Manfaat Penelitian.....	4
1.5 Metode Penelitian.....	5
1.5.1 Metode Studi Literatur	5
1.5.2 Metode Data Mining	5

1.5.3	Metode Pengujian.....	7
1.6	Sistematika Penulisan.....	7
BAB II LANDASAN TEORI		9
2.1	Kajian Pustaka.....	9
2.2	Kaggle	16
2.3	Sistem Rekomendasi	16
2.4	Content Based Filtering (CBF).....	17
2.5	Collaborative Filtering (CF).....	18
2.5.1	User-Based Collaborative Filtering (UB-CF).....	19
2.5.2	Item Based Collaborative Filtering (IB-CF)	20
2.6	Pearson Correlation Coefficient	21
2.7	Cosine Similarity	23
2.8	K-Nearest Neighbor	24
2.9	Root Mean Square Error (RMSE).....	25
2.10	Python.....	26
2.11	Google Collab.....	27
2.12	Surprise	27
BAB III METODE PENELITIAN		28
3.1	Alur Penelitian.....	28
3.2	Pengambilan Data.....	29

3.3	Preprocessing Data (Pembersihan Data)	29
3.4	Analisis Data	30
3.5	Data Training dan Data Test	30
3.6	Implementasi Algoritma Cosine dan Pearson Similarity	30
3.7	Perhitungan Manual Cosine dan Pearson Similarity	31
3.7.1	Cosine Similarity	31
3.7.2	Pearson Similarity	41
3.8	Tuning Parameter	51
3.9	Alat dan Bahan Penelitian	51
3.9.1	Kebutuhan Perangkat Keras	51
3.9.2	Kebutuhan Perangkat Lunak	52
3.10	Hasil Evaluasi	53
BAB IV HASIL DAN PEMBAHASAN		54
4.1	Persiapan Data	54
4.1.1	Pengumpulan data	54
4.2	Preprocessing Data	55
4.2.1	Preprocessing Dataset BookCrossing	56
4.2.2	Preprocessing Dataset Good Books	58
4.3	Analisis Data	59
4.4	Implementasi dan Pengujian Algoritma	60



4.4.1	Pengujian Pada Dataset Good Books.....	62
4.4.2	Pengujian Pada Dataset BookCrossing.....	64
4.5	Tuning Parameter	65
4.5.1	Grid Search CV Dataset Good Books.....	66
4.5.2	Grid Search CV Dataset Book Crossing.....	68
4.6	Analisis hasil model Collaborative Filtering.....	70
4.6.1	Dataset Good Books.....	70
4.6.2	Dataset Book Crossing.....	73
BAB V KESIMPULAN DAN SARAN.....		77
5.1	Kesimpulan.....	77
5.2	Saran.....	78
DAFTAR PUSTAKA.....		lxxix

DAFTAR GAMBAR

Gambar 2. 1 Konsep User-based Collaborative filtering	20
Gambar 2. 2 Konsep Item-based Collaborative filtering.....	21
Gambar 3. 1 Diagram Alur Tahapan Penelitian	28
Gambar 4. 1 Dataset Ratings - Goodbooks10k-updated	54
Gambar 4. 2 Dataset Ratings – Book Crossing.....	55
Gambar 4. 3 Pembuatan fungsi operasi dataframe.....	55
Gambar 4. 4 Pemanggilan fungsi yang telah dibuat pada google collab.....	56
Gambar 4. 5 Pengecekan karakter pada dataset books.csv BookCrossing.....	56
Gambar 4. 6 Proses Filter dataset BookCrossing.....	57
Gambar 4. 7 Proses pengurangan dimensi pada dataset rating BookCrossing....	57
Gambar 4. 8 Penyimpanan Dataset rating dan hasil preprocessing BookCrossing	57
Gambar 4. 9 Dataset Good Books sebelum dilakukan preprocessing.....	58
Gambar 4. 10 Hasil pengurangan dimensi Dataset Good Books	58
Gambar 4. 11 Distribusi rating BookCrossing	59
Gambar 4. 12 Jumlah rating, buku dan user pada dataset BookCrossing	60
Gambar 4. 13 Distribusi rating Good Books	60
Gambar 4. 14 Jumlah rating, buku dan user pada dataset Good Books	60
Gambar 4. 15 Sintak fungsi generate_model_dict	61
Gambar 4. 16 Sintak fungsi draw_model_results.....	61
Gambar 4. 17 Sintak pengujian dan hasil pengujian model dataset Good Books	62
Gambar 4. 18 Visualisasi hasil pengujian model dataset Good Books	63

Gambar 4. 19 Sintak pengujian dan hasil pengujian model dataset BookCrossing	64
Gambar 4. 20 Visualisasi hasil pegujian model dataset BookCrossing	65
Gambar 4. 21 Sintak pengujian Grid Search CV.....	66
Gambar 4. 22 Hasil dari Uji Nilai K 1-10	66
Gambar 4. 23 Hasil keseluruhan Grid Search CV 1-100	67
Gambar 4. 24 Sintak pengujian Grid Search CV.....	68
Gambar 4. 25 Hasil keseluruhan Grid Search CV 1-100	69
Gambar 4. 26 Sintaks pengujian ulang dengan parameter terbaik	70
Gambar 4. 27 Sintaks Pembuatan tabel hasil model	71
Gambar 4. 28 Sintaks dan hasil dari distribusi rating sebenarnya dengan hasil prediksi.....	71
Gambar 4. 29 Sintaks dan hasil prediksi untuk spesifik user	72
Gambar 4. 30 Sintaks dan Hasil perbandingan distribusi antara rating sebenarnya dan prediksi rating.....	72
Gambar 4. 31 Sintaks pengujian ulang dengan parameter terbaik.....	73
Gambar 4. 32 Sintaks Pembuatan tabel hasil model	73
Gambar 4. 33 Sintaks dan hasil dari distribusi rating sebenarnya dengan hasil prediksi.....	74
Gambar 4. 34 Sintaks dan hasil prediksi untuk spesifik user	74
Gambar 4. 35 Sintaks dan Hasil perbandingan distribusi antara rating sebenarnya dan prediksi rating.....	75

DAFTAR TABEL

Tabel 2. 1 Tabel Perbandingan.....	13
Tabel 3. 1 Sampel Data Rating.....	32
Tabel 3. 2 Hasil perhitungan Similarity setiap item.....	35
Tabel 3. 3 hasil evaluasi RMSE dengan nilai $K = 3$	40
Tabel 3. 4 hasil evaluasi RMSE dengan nilai $K = 4$	40
Tabel 3. 5 Sampel Data Rating.....	42
Tabel 3. 6 Normalisasi Rating.....	42
Tabel 3. 7 Hasil perhitungan Similarity setiap item.....	45
Tabel 3. 8 hasil evaluasi RMSE dengan nilai $K = 3$	49
Tabel 3. 9 hasil evaluasi RMSE dengan nilai $K = 4$	50
Tabel 3. 10 Kebutuhan Perangkat Keras.....	51
Tabel 3. 11 Kebutuhan Perangkat Lunak.....	52
Tabel 4. 1 Analisis dataset.....	59
Tabel 4. 2 Hasil evaluasi RMSE Good Books.....	67
Tabel 4. 3 Hasil evaluasi RMSE Good Books.....	69

INTISARI

Sistem rekomendasi menjadi populer di banyak hal yang dijual di internet seperti buku, film, artikel penelitian, media promosi seperti iklan atau ads dan banyak lagi, salah satu yang paling populer adalah sistem rekomendasi pilihan film pada layanan *Netflix*. *Netflix* membuat rekomendasi berdasarkan film yang telah dilihat pelanggan sebelumnya. Pada tahun 2006 *Netflix* mengadakan kompetisi kepada orang pertama yang dapat meningkatkan sistem rekomendasi sebesar 10% dengan total hadiah 1 juta *dollar*. Dapat kita sadari bahwa peningkatan kinerja sistem rekomendasi memiliki kompleksitas dan merupakan hal yang berharga.

Pada penelitian ini terdapat dua algoritma dalam mencari tingkat kesamaan antar *item* yaitu *Cosine similarity* dan *Pearson Similarity* yang akan diterapkan pada data *Good Books* dan *Book Crossing* yang di implementasi menggunakan *Google Colab* dengan menggunakan *library Surprise*. Setelah melewati pengujian, kemudian setiap algoritma akan diuji Kembali dengan *tuning hyperparameter* agar dapat mendapatkan hasil nilai *k* terbaik. Penelitian ini bertujuan untuk mengetahui algoritma kesamaan yang memiliki hasil evaluasi terbaik dengan pengujian evaluasi menggunakan metode *root mean square error* (RMSE).

Dari hasil pengujian didapatkan hasil evaluasi terbaik adalah *pearson* daripada *cosine* pada dataset *Good Books* didapatkan nilai RMSE sebesar 0.919560 dan pada dataset *Book Crossing* sebesar 1.805998. Pada pengujian *tuning hyperparameter* cukup memberikan perubahan signifikan mendekati 0.1. pada dataset *Good Books* menjadi 0.894753 dan pada dataset *Book Crossing* menjadi 1.720001.

Kata Kunci: *Sistem Rekomendasi, Item Based Collaborative Filtering, Perbandingan Similaritas, tuning nilai k, hyperparameter, surprise, GridSearchCV*

ABSTRACT

Recommendation systems are becoming popular in many things that are sold on the internet such as books, movies, research articles, promotional media such as advertisements or ads and many more, one of the most popular is the movie selection recommendation system on the Netflix service. Netflix makes recommendations based on movies that customers have seen late in the past. In 2006 Netflix held a competition to the first person who could increase the recommendation system by 10% with a total prize of 1 million dollars. We can recognize that improving the performance of recommendation systems is complex and valuable.

In this study, there are two algorithms for finding the level of similarity between items, namely Cosine similarity and Pearson Similarity which will be applied to Good Books and Book Crossing data which are implemented using Google Colab using the Surprise library. After passing the test, then each algorithm will be tested again with hyperparameter tuning in order to get the best k value results. This study aims to determine the similarity algorithm that has the best evaluation results by evaluating the test using the root mean square error (RMSE) method.

From the test results, the best evaluation results are pearson rather than cosine in the Good Books dataset, the RMSE value is 0.919560 and the Book Crossing dataset is 1.805998. In the hyperparameter tuning test, it is enough to give a significant change close to 0.1. on the Good Books dataset it becomes 0.894753 and on the Book Crossing dataset it becomes 1.720001.

Keywords : Recommendation System, Item Based Collaborative Filtering, Similarity Comparison, k value tuning, hyperparameter, surprise, GridSearchCV