JTRISTE, Vol.13, No.2, Oktober 2026, pp. 45~57
ISSN: **2355-3677 | e-ISSN : 2460-8548**

Model *Machine Learning* Untuk Klasifikasi Kemiripan Komik Menggunakan *Jaccard Similarity* Dan *Logistic Regression*

Richie Theodore[1], Moh. Sofyan S. Thayf[2*] , Zaenab Pontoh[3]
[1,2,3]Sistem Informasi, STMIK KHARISMA Makassar
e-mail: [1]richietheodore_23@kharisma.ac.id, [2]sofyan.thayf@kharisma.ac.id,
[3]zaenabp@kharisma.ac.id

Abstrak

TeenToon merupakan platform komik digital berbasis mobile yang menyajikan komik dengan nilai-nilai Kristiani. Seiring bertambahnya jumlah komik yang tersedia, platform ini menghadapi permasalahan information overload, sehingga pengguna mengalami kesulitan dalam menemukan komik yang sesuai dengan preferensi mereka. Penelitian ini bertujuan melakukan klasifikasi kemiripan komik untuk sistem rekomendasi pada aplikasi TeenToon menggunakan pendekatan Content-Based Filtering dengan menerapkan Jaccard Similarity dan Logistic Regression. Kemiripan antar komik dihitung berdasarkan atribut genre dan tag, kemudian hasil rekomendasi disajikan dalam bentuk 5 rekomendasi teratas. Penelitian menggunakan 25 judul komik yang dikombinasikan sehingga menghasilkan 300 pasangan data. Dari total pasangan, dataset dibagi menjadi 248 pasangan data latih dan 52 pasangan data uji. Setiap pasangan diberi label oleh tiga validator berdasarkan penilaian subjektif terhadap tingkat kemiripan komik dengan mempertimbangkan alur cerita serta kesesuaian atribut genre dan tag. Label akhir ditentukan menggunakan metode majority voting, sedangkan nilai Jaccard Similarity digunakan sebagai fitur masukan pada model klasifikasi Logistic Regression. Model yang dihasilkan kemudian diimplementasikan pada halaman detail komik di aplikasi TeenToon. Hasil pengujian menunjukkan bahwa model Logistic Regression mencapai akurasi sebesar 88,46%, presisi sebesar 71,43%, recall sebesar 100%, F1-score sebesar 83,33%, ROC-AUC sebesar 0,982. Temuan ini menunjukkan bahwa Logistic Regression mampu diterapkan untuk membangun sistem rekomendasi berbasis kemiripan atribut pada aplikasi TeenToon.

Kata kunci: *Content-Based Filtering, Jaccard Similarity, Logistic Regression, Sistem Rekomendasi, Komik Digital.*

Abstract

TeenToon is a mobile-based digital comic platform that provides digital comics that incorporate Christian values. As the number of available comics continues to grow, the platform faces the challenge of information overload, making it difficult for users to discover comics that match their preferences. This study aimed to classify comic similarity for a recommendation system on TeenToon using a Content-Based Filtering approach by applying Jaccard Similarity and Logistic Regression. The similarity between comics was calculated based on genre and tag attributes, and the recommendations were presented as the Top 5 recommendations. The study utilized 25 comic titles, which were paired to generate 300 data pairs. From the total pairs, the dataset was split into 248 training pairs and 52 test pairs. Each pair was labeled by three validators based on subjective assessments of comic similarity, considering both the storyline and the similarity of genre and tag attributes. The final labels were determined using majority voting. The Jaccard Similarity score was then calculated and used as an input feature for the Logistic Regression

[*] Corresponding author : Moh. Sofyan S. Thayf(sofyan.thayf@kharisma.ac.id)

***Diterima** : Agustus, 2026* ***Disetujui** : Oktober, 2026* ***Dipublikasikan** : Oktober, 2026*

classification model, which was subsequently implemented on the comic detail page of the TeenToon application. The results showed that the Logistic Regression model achieved an accuracy of 88.46%, a precision of 71.43%, a recall of 100%, an F1-score of 83,33%, and a ROC-AUC of 0.982. These findings indicate that Logistic Regression is capable of being applied to develop an attribute-similarity-based recommendation system for the TeenToon application.

Keywords: *Content-Based Filtering, Jaccard Similarity, Logistic Regression, Recommendation System, Digital Comics.*

1. Pendahuluan

Transformasi digital telah mengubah cara masyarakat mengakses konten hiburan, termasuk komik, yang ditandai dengan munculnya platform komik digital sejak diperkenalkannya Line Webtoon di Indonesia pada tahun 2014. Di balik kemudahan akses yang ditawarkan, transformasi ini turut membawa tantangan terhadap pola konsumsi informasi generasi muda. Remaja cenderung mengalihkan waktu luangnya dari aktivitas literasi ke arah hiburan seperti game atau media sosial [1], termasuk rendahnya minat terhadap literasi keagamaan yang dianggap membosankan [2]. Kondisi ini mendorong perlunya inovasi media alternatif yang mengemas nilai-nilai spiritual ke dalam format visual yang modern dan relevan bagi remaja.

TeenToon dikembangkan sebagai respons atas kebutuhan tersebut, yaitu platform komik digital berbasis *mobile* yang mengombinasikan berbagai genre komik dengan unsur Kristiani, dengan remaja sebagai target utama pengguna. Seiring bertambahnya jumlah konten komik pada TeenToon, muncul permasalahan *information overload*, di mana pengguna kesulitan menemukan komik yang sesuai dengan preferensi genre dan cerita [3], sementara fitur rekomendasi yang telah direncanakan belum diimplementasikan pada aplikasi.

Untuk mengatasi permasalahan tersebut, mengingat *Collaborative Filtering* memerlukan data historis interaksi antarpengguna [3] yang belum memadai pada Teentoon akibat keterbatasan jumlah pengguna, sehingga sistem menggunakan metode *Content-Based Filtering*, metode ini menghasilkan rekomendasi berdasarkan kemiripan atribut konten tanpa bergantung pada data pengguna lain [4]. Kemiripan antar komik diukur menggunakan *Jaccard Similarity* berdasarkan atribut genre dan *tag*, metode ini efektif pada data teks pendek berbasis kategorikal [5], [6] dibandingkan *Cosine Similarity* yang lebih unggul pada teks naratif panjang seperti sinopsis [7], maupun *Dice Similarity* yang cenderung menghasilkan nilai kemiripan lebih tinggi [8], hal itu karena metode ini tidak mengurangi nilai irisan pada penyebutnya. Skor kemiripan yang dihasilkan *Jaccard Similarity* kemudian diklasifikasikan menggunakan *Logistic Regression* untuk menentukan kategori "mirip" atau "tidak mirip" secara lebih akurat.

Penerapan pendekatan dan algoritma dalam sistem rekomendasi telah banyak dilakukan pada berbagai topik. Alana dan Hartanto [9] menerapkan *Cosine Similarity* pada rekomendasi komik berdasarkan sinopsis dengan *precision* 94,86%, sementara Irwan et al [10] menerapkannya pada rekomendasi anime dengan *Mean Average Precision* 0,4167. Fahrizal & Hasugian [11] menggunakan KNN dan *Jaccard Similarity* pada rekomendasi TV series dengan *precision* dan *recall* mencapai 1,0 pada genre spesifik.

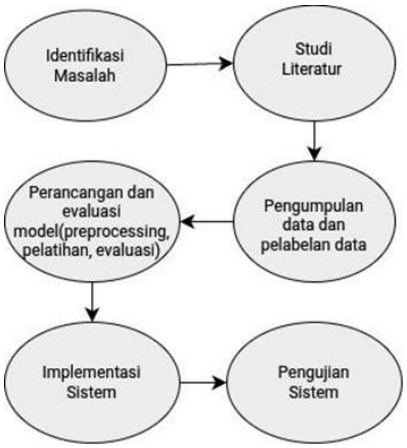
Namun, belum ditemukan penelitian yang menggabungkan *Jaccard Similarity* dengan *Logistic Regression* secara khusus pada domain rekomendasi komik digital. Kebaruan penelitian ini terletak pada objek penelitian dan skemanya, di mana penelitian berfokus pada penentuan label kemiripan (mirip atau tidak) untuk menghasilkan rekomendasi kemiripan komik.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem rekomendasi komik digital pada aplikasi TeenToon menggunakan pendekatan *Content-Based Filtering*, menerapkan algoritma *Jaccard Similarity* dan *Logistic Regression* untuk menghasilkan klasifikasi rekomendasi yang relevan berdasarkan kemiripan konten komik yang sedang diakses pengguna, serta menguji fungsionalitas sistem rekomendasi pada aplikasi TeenToon menggunakan metode *Black Box Testing* untuk memastikan sistem berjalan sesuai skenario yang diharapkan.

2. Metode Penelitian

2.1 Tahapan Penelitian

Penelitian ini dirancang menggunakan pendekatan kuantitatif untuk membangun sistem rekomendasi berbasis *Content-Based Filtering* pada aplikasi TeenToon. Tahapan penelitian yang dilakukan pada penelitian ini melalui enam tahap seperti Gambar 1 sebagai berikut:



Gambar 1. Tahapan Penelitian

Penelitian ini dirancang menggunakan pendekatan kuantitatif untuk membangun sistem rekomendasi berbasis *Content-Based Filtering* pada aplikasi TeenToon, melalui enam tahap: (1) identifikasi masalah terkait *information overload* pada platform komik TeenToon; (2) studi literatur mengenai sistem rekomendasi, *Content-Based Filtering*, *Jaccard Similarity*, dan *Logistic Regression*; (3) pengumpulan dan pelabelan data atribut komik dari Firebase; (4) perancangan, pelatihan, dan evaluasi model *Logistic Regression* menggunakan *Confusion Matrix* dengan metrik *Accuracy* dan *Precision*; (5) implementasi model ke *backend* Python yang terhubung dengan *frontend* Android (Kotlin); serta (6) pengujian sistem menggunakan metode *black box testing*, metode yang bertujuan menilai fungsionalitas sistem tanpa perlu tahu struktur kode aplikasi [12], melalui empat skenario kombinasi kemiripan *genre* dan *tag*.

2.2 Pengumpulan Data dan Pelabelan Data

Data penelitian bersumber dari atribut komik (ID, judul, genre, dan *tag*) yang diekstrak dari Firebase Realtime Database yang menggunakan data riil yang merepresentasikan konten komik yang aktif pada aplikasi TeenToon. Proses pengambilan data dilakukan pada periode Februari hingga Juli 2026. Penentuan 25 judul komik dilakukan berdasarkan kriteria inklusi berikut: (1) komik telah rilis secara resmi pada periode tersebut, (2) komik memiliki kelengkapan atribut yang wajib ada untuk perhitungan kemiripan. Untuk mengukur kemiripan antar-komik, 25 komik atribut genre dan *tag* dikombinasikan sehingga menghasilkan 300 pasangan komik unik dengan rumus dibawah ini [13]:

$$C(n,k) = \frac{n!}{(n-k)!k!} \quad \text{.....(1)}$$

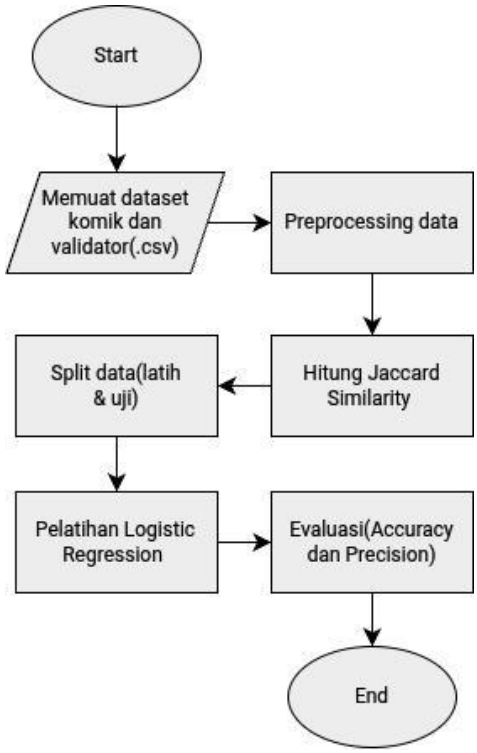
Di mana n adalah jumlah data komik (n = 25) dan k adalah jumlah komik dalam satu pasangan (k = 2) (1). Ukuran dataset tersebut merupakan salah satu keterbatasan penelitian. Oleh karena itu, *dataset* sebanyak 300 pasangan komik dibagi berdasarkan judul komik dengan proporsi 80% data latih (248 sampel) dan 20% data uji (52 sampel) untuk mencegah kebocoran data (*data leakage*). Berdasarkan perhitungan C(25, 2), selanjutnya proses pelabelan data (*ground truth*) dilakukan menggunakan lembar kerja Excel yang diserahkan kepada tiga orang validator (terdiri dari kreator dan pengguna aktif TeenToon). Validator memberikan penilaian biner pada setiap pasangan komik di mana label 1: Menandakan pasangan komik bernilai mirip sedangkan label 0: Menandakan pasangan komik bernilai tidak mirip.

Proses pelabelan tersebut dilakukan dengan kriteria operasional kemiripan yang mengharuskan adanya kesamaan alur cerita yang didukung oleh salah satu dari dua kondisi berikut: (1) pasangan komik yang memiliki satu atau lebih genre yang sama serta minimal satu *tag* yang sama, atau (2) pasangan komik yang memiliki genre berbeda namun memiliki minimal

dua tag yang sama. Apabila kriteria tidak terpenuhi, pasangan komik dikategorikan sebagai tidak mirip (label 0). Tingkat kesepakatan validator dihitung menggunakan Fleiss' Kappa. Berdasarkan hasil pelabelan yang dilakukan oleh tiga validator, diperoleh nilai Fleiss' Kappa sebesar 0.7030, yang mengindikasikan tingkat kesepakatan substansial (*substantial agreement*) berdasarkan rentang kriteria 0,61 - 0,80 menurut skala Landis dan Koch [14] di antara validator. Selanjutnya, keputusan akhir label *ground truth* ditentukan menggunakan *majority vote* (suara terbanyak) untuk menjaga objektivitas penilaian data uji.

2.1 Perancangan dan Evaluasi Model *Machine Learning*

Proses perancangan dan evaluasi model *Machine Learning* melalui beberapa tahapan seperti Gambar 2 sebagai berikut:



Gambar 2. Flowchart alur pelatihan dan evaluasi model

1. Memuat *dataset* komik: Melakukan proses muat *dataset* komik yang telah melewati proses pelabelan oleh validator, *dataset* komik ini dalam bentuk .csv lalu dimuat dengan pustaka Pandas sebagai df(*data frame*).
2. *Preprocessing* Teks: Melakukan pemecahan teks (*tokenization*) menggunakan fungsi *split*, pembersihan spasi (*stripping*) menggunakan fungsi *strip*, penyeragaman huruf (*case folding*) yang berfungsi untuk mencegah *case sensitivity* [15] menggunakan fungsi *lower*, serta transformasi ke bentuk himpunan (*set*) sebagai representasi atribut untuk perhitungan *Jaccard Similarity*.
3. Kalkulasi Kemiripan (*Jaccard Similarity*): Mengubah genre dan *tag* dari pasangan komik A dan B menjadi bentuk himpunan, kemudian menghitung nilai kemiripannya menggunakan rumus [16]:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad \text{.....(2)}$$

Keterangan (2):
$J(A,B)$:Nilai kemiripan *Jaccard Similarity* antara komik A dan komik B
$|A \cap B|$: Jumlah elemen genre atau *tag* yang sama (irisan) antara komik A dan Komik B yang diwakili oleh operasi *len(set_a & set_b)* pada Python
$|A \cup B|$: Jumlah elemen genre atau *tag* keseluruhan (gabungan) dari komik A dan komik B yang diwakili oleh operasi *len(set_a | set_b)* pada Python

4. Pembagian Data (*Data Splitting*): *Dataset* sebanyak 300 pasangan komik dibagi dengan rasio 80% untuk data latih (*training set*) dan 20% untuk data uji (*testing set*).

5. Pelatihan Model: Model *Logistic Regression* dilatih menggunakan data latih untuk mempelajari pola kemiripan untuk memprediksi probabilitas berdasarkan skor *Jaccard Similarity* genre (x_1) dan *tag* (x_2). proses kalkulasi probabilitas ini menggunaka fungsi sigmoid sebagai berikut [17]:

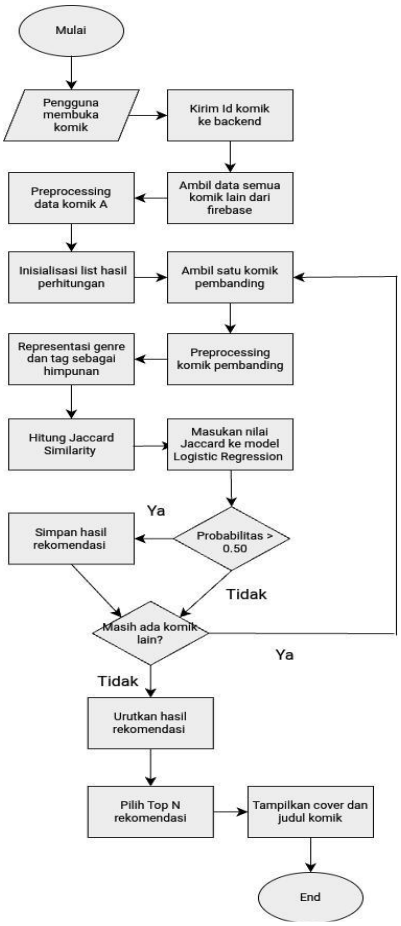
$$P(X) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2}} \quad \text{..... (3)}$$

Keterangan (3):
$P(X)$: Probabilitas hasil prediksi dengan nilai rentang antara 0 sampai 1
e : bilangan konstanta Euler, yang nilainya kurang lebih 2,718.
β_0 (*Intercept*) : konstanta atau titik potong pada sumbu y
β_1, β_2 (*Weight*) : koefisien regresi untuk fitur *Jaccard Similarity* genre dan *Jaccard Similarity tag*
X_1, X_2 : Nilai *input* dari *Jaccard Similarity* genre dan *Jaccard Similarity tag*.
Model yang telah terlatih kemudian disimpan dalam format *served model* (.pkl) untuk diintegrasikan pada *backend*

6. Evaluasi Performa: Pengujian dilakukan menggunakan data uji (20%) yang direpresentasikan melalui *Confusion Matrix* untuk memperoleh nilai *Accuracy* dan *Precision* sebagai metrik utama keberhasilan model. Penetapan batas klasifikasi biner menggunakan nilai ambang batas (*threshold*) probabilitas sebesar p > 0,50 [17].

2.2 Implementasi Sistem

Model *Logistic Regression* (.pkl) diintegrasikan ke dalam *backend* berbasis Python dengan *framework* FastAPI. *Backend* ini berfungsi menyediakan API *endpoint* yang terhubung dengan aplikasi *mobile* TeenToon pada Android Studio (Kotlin). Alur proses sistem rekomendasi dapat dilihat pada Gambar 3



Gambar 3. Flowchart alur sistem rekomendasi

Berdasarkan Gambar 3, proses alur kerja sistemnya dimulai ketika pengguna membuka halaman detail salah satu komik, *frontend* Android secara otomatis mengirimkan ID komik tersebut ke *backend*, selanjutnya *backend* mengambil data komik pembanding dari Firebase, melakukan *preprocessing*, dan menghitung nilai *Jaccard Similarity*, lalu model *Logistic Regression* memprediksi probabilitas kemiripan. Pasangan komik dengan nilai probabilitas > 0.50 dinyatakan mirip, langkah terakhir, sistem mengurutkan hasil komik yang mirip secara menurun (*descending*) berdasarkan skor probabilitas terkini dan menyajikan Top-5 komik rekomendasi pada layar aplikasi pengguna.

2.3 Pengujian Sistem

Pengujian sistem dilakukan menggunakan metode *black box testing* untuk menguji fitur rekomendasi komik pada aplikasi TeenToon. Rancangan pengujian mengacu pada struktur tabel skenario pengujian yang umum digunakan dalam penelitian sejenis, sebagaimana diterapkan oleh [18] pengujian aplikasi monitoring pemberian obat filariasis berbasis Android. Pengujian dilakukan dengan membuka komik tertentu pada halaman detail komik, kemudian mengamati kesesuaian daftar rekomendasi Top 5 yang ditampilkan berdasarkan empat skenario tingkat kemiripan atribut genre dan *tag* terhadap komik lain dalam database. Rancangan skenario pengujian disajikan pada Tabel 1

Tabel 1. Rancangan Pengujian Sistem

Kode Test Case	Skenario	Kondisi *Input*	Hasil yang diharapkan
TC001	Genre mirip, *tag* mirip	Membuka komik Ken's Journey	Komik dengan genre dan *tag* mirip seperti Kisah si Botak, Life Boys muncul di *Top* 5 *recommendation*
TC002	Genre tidak mirip, *tag* mirip	Membuka komik Back to School	Komik dengan genre tidak mirip namun *tag* mirip seperti Our Team, Bestie tetap muncul di *Top* 5 *recommendation*
TC003	Genre mirip, *tag* tidak mirip	Membuka komik Food Bible	Komik dengan genre mirip namun *tag* tidak mirip seperti Buble Faith, Carlos tetap muncul di *Top* 5 *recommendation*
TC004	Genre tidak mirip, *tag* tidak mirip	Membuka komik Dating Challenge	Komik yang genre & *tag*nya sepenuhnya tidak mirip seperti Daily Comic Artist, Pride and Writer tidak muncul di *Top* 5 *recommendation*

1. Hasil dan Pembahasan

3.1 Hasil Pengolahan Data dan Pelabelan (*Ground Truth*)

Dataset penelitian diekstrak dari Firebase Realtime Database berupa atribut 25 judul komik (judul, genre, dan *tag*). Menggunakan rumus kombinasi C(25, 2), terbentuk 300 pasangan komik unik. Pelabelan *ground truth* oleh tiga orang validator melalui *majority vote* menghasilkan distribusi label biner yang disajikan pada Tabel 2.

Tabel 2. Distribusi Label

Label	Keterangan	Jumlah	Persentase
0	Tidak Mirip	214	71.33%
1	Mirip	86	28.67%

Setelah tahap *preprocessing* (*tokenization*, *stripping*, *case folding*, dan konversi ke tipe set), dilakukan kalkulasi koefisien kemiripan menggunakan *Jaccard Similarity* secara terpisah

untuk atribut genre (jaccard_genre) dan *tag* (jaccard_tag). Sampel hasil kalkulasi yang menjadi fitur masukan (x_1 dan x_2). disajikan pada Tabel 3.

Tabel 3. Distribusi Hasil *Jaccard Similarity*

No	Judul A	Judul B	jaccard_genre	jaccard_tag	Label
1	Ken's Journey	Buble Faith	0.500000	0.333333	1
2	Ken's Journey	The Girls Life	1.000000	0.250000	1
3	Ken's Journey	Life Boys	1.000000	0.666667	1
4	Ken's Journey	My Horror Story	0.333333	0.333333	0
5	Ken's Journey	Bestie	0.333333	0.250000	0
6	Ken's Journey	Daily Bread Bakery	1.000000	0.250000	1
...	...	...	...	...	...

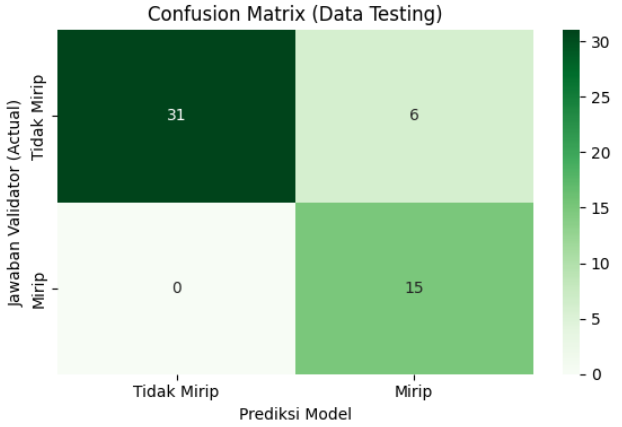
3.2 Hasil Pelatihan dan Evaluasi Model *Logistic Regression*

Model *Logistic Regression* dilatih menggunakan parameter *class_weight='balanced'* guna mengatasi ketidakseimbangan kelas (*class imbalance*). Bobot parameter optimal yang dihasilkan dari proses pelatihan disajikan pada Tabel 4.

Tabel 4. Nilai Bobot Hasil Pelatihan Model

Parameter	Representasi Variabel	Nilai
Intercept/Bias	β_0	-1.965434
Bobot *Jaccard Similarity* genre	β_1	2.307600
Bobot *Jaccard Similarity tag*	β_2	4.352532

Setelah itu pengujian performa model pada 52 data uji menggunakan *Confusion Matrix* menghasilkan tingkat keberhasilan klasifikasi yang disajikan pada Gambar 4.



Gambar 4. Hasil *Confusion Matrix*

Berikut rumus *accuracy*, *precision, recall,* F1-*score* yang digunakan sebagai metrik evaluasi:

Accuracy [19]: $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$

Precision [19], [20]: $Precision = \frac{\mathrm{TP}}{(\mathrm{TP+FP})}$

Recall [19] : $Recall = \frac{\mathrm{TP}}{(\mathrm{TP+FN})}$

F1-*Score* [19] : $F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$

ROC-AUC merupakan metrik evaluasi yang digunakan untuk menilai kemampuan model dalam proses membedakan kelas positif dan negatif [19].

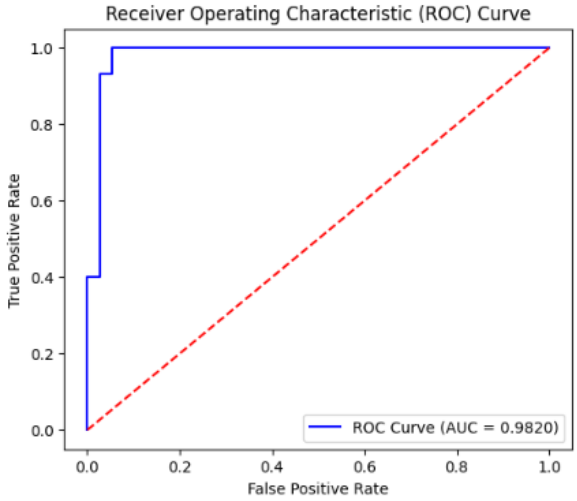
Berdasarkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN) pada *Confusion Matrix* dimasukkan pada rumus *Accuracy*, *Precision, Recall,* F1-

Score, serta melakukan evaluasi ROC-AUC maka didapatkan hasil yang dapat dilihat pada Tabel 5:

Tabel 5. Hasil *Accuracy*, *Precision, Recall, F1-Score, ROC-AUC*

Metrik Evaluasi	Nilai Skor	Persentase
Accuracy	0.884615	88.46%
Precision	0.714286	71.43%
Recall	1.000000	100.00%
F1-*Score*	0.833333	83.33%
ROC-AUC	0.981982	98.20%

Hasil evaluasi performa model klasifikasi menggunakan kurva *Receiver Operating Characteristic* (ROC) dapat dilihat pada Gambar 5. Berdasarkan grafik pada Gambar 5, model *Logistic Regression* menghasilkan *Area Under Curve* (AUC) sebesar 0.982.



Gambar 5. Hasil kurva ROC

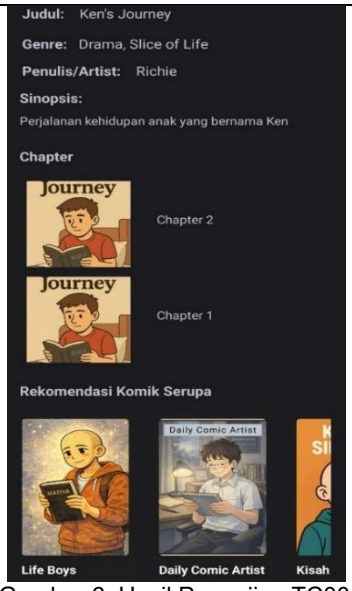
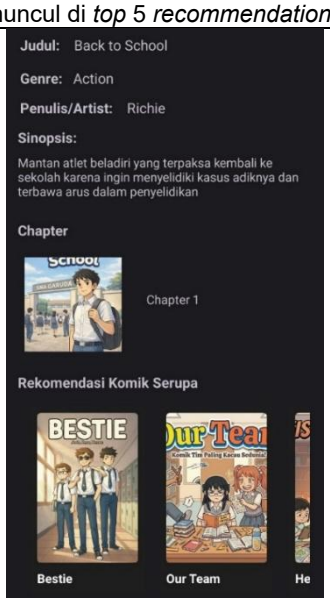
3.3 Implementasi Fitur Rekomendasi pada Aplikasi TeenToon

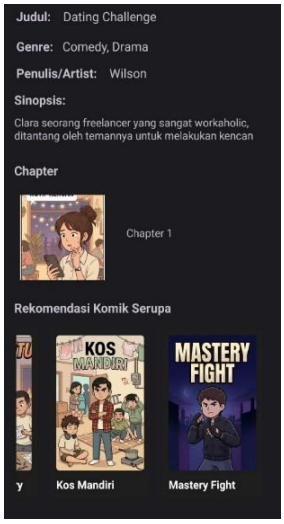
Model terlatih disimpan ke dalam format .pkl dan di-*deploy* pada *backend* berbasis Python (FastAPI). *Backend* menerima masukan ID komik, menghitung nilai probabilitas kemiripan dengan fungsi *.predict_proba(),* menyaring komik dengan nilai probabilitas > 0,50, dan mengembalikan daftar Top 5 komik rekomendasi dalam format JSON secara *descending*. Pada sisi *frontend*, aplikasi TeenToon berbasis Android (Kotlin) secara otomatis meminta data rekomendasi saat pengguna membuka halaman detail komik.

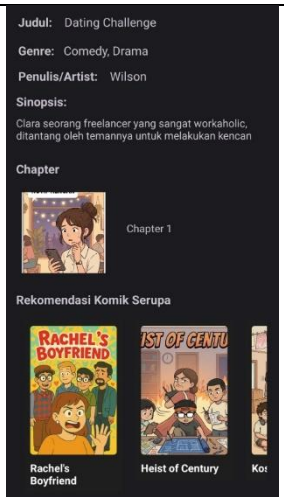
3.4 Pengujian Sistem

Berdasarkan rancangan pengujian pada Tabel 1, dilakukan pengujian *black box testing* terhadap fitur rekomendasi komik pada aplikasi TeenToon. Berikut hasil pengujian sistem seperti pada Tabel 6 dibawah ini

Tabel 6. Hasil Pengujian *Black Box testing*

Kode Test Case	Skenario	Kondisi *input*	Hasil yang diharapkan	Hasil aktual	*Output* hasil
TC001	Genre mirip, *tag* mirip	Membuka komik Ken's Journey	Komik dengan genre dan *tag* mirip seperti Kisah si Botak, Life Boys muncul di *top* 5 *recommendation*	 Gambar 6. Hasil Pengujian TC001 Berdasarkan Gambar 6, Komik Kisah si Botak dan Life Boys muncul di *top* 5 *recommendation*	Berhasil
TC002	Genre tidak mirip, *tag* mirip	Membuka komik Back to School	Komik dengan genre tidak mirip namun *tag* mirip seperti Our Team, Bestie tetap muncul di *top* 5 *recommendation*	 Gambar 7. Hasil Pengujian TC002 Berdasarkan Gambar 7, komik Our Team dan Bestie tetap muncul di *top* 5 *recommendation*	Berhasil

TC003	Genre mirip, *tag* tidak mirip	Membuka komik Food Bible	Komik dengan genre mirip namun *tag* tidak mirip seperti Buble Faith, Carlos tetap muncul di *top* 5 *recommendation*	 Gambar 8. Hasil Pengujian TC003 Berdasarkan Gambar 8, komik Buble Faith, Carlos tetap muncul di *top* 5 *recommendation*	Berhasil
TC004	Genre tidak mirip, *tag* tidak mirip	Membuka komik Dating Challenge	Komik yang genre dan *tag*nya sepenuhnya tidak mirip seperti Daily Comic Artist, Pride and Writer tidak muncul di *top* 5 *recommendation*	 Gambar 9. Hasil Pengujian TC004	Berhasil

				 Gambar 10. Hasil Pengujian TC004 Berdasarkan Gambar 9 dan 10, komik Daily Comic Artist dan Pride and Writer tidak muncul di *top* 5 *recommendation*	

3.5 Pembahasan

Berdasarkan hasil pengujian pada Tabel 5, model *Logistic Regression* berhasil mencapai tingkat *Accuracy* sebesar 88,46%, *Precision* sebesar 71,43%, *Recall* sebesar 100%, F1-*Score* sebesar 83,33%, dan nilai ROC-AUC sebesar 0,982. Tingginya nilai *accuracy* menunjukkan proporsi keseluruhan prediksi yang benar, di mana model mampu mengenali pasangan komik yang relevan maupun tidak, sementara *precision* menunjukkan ketepatan rekomendasi yang diberikan. Hal ini berkaitan erat dengan *false positive* yang terjadi ketika model memprediksi komik memiliki kemiripan dan merekomendasikan ke pengguna padahal secara fakta komik tersebut tidak relevan, sebaliknya *false negative* terjadi ketika model gagal mengenali komik yang sebenarnya ada kemiripan, sehingga terjadi kehilangan peluang menyajikan komik relevan.

Nilai *recall* menunjukkan kemampuan model dalam menemukan komik yang relevan. F1-*score* merepresentasikan keseimbangan antara *precision* dan *recall*. Selain itu, nilai ROC-AUC sebesar 0,982 menunjukkan bahwa model mampu mendefinisikan batas kemiripan. Namun, evaluasi performa ini perlu diinterpretasikan dalam konteks ukuran dataset uji yang relatif kecil sebagai keterbatasan penelitian. Untuk memvalidasi performa pada distribusi data yang tidak seimbang, evaluasi mengacu pada prinsip [17]. Pada *dataset* uji (52 data), kelas mayoritas ('Tidak Mirip') berjumlah 37 sampel, sehingga pemodelan dasar (*trivial null classifier*) hanya menghasilkan akurasi dasar sebesar 71,15%. Perolehan akurasi model yang melampaui akurasi dasar dengan selisih 17,31% membuktikan bahwa model mampu mempelajari pola fitur *Jaccard Similarity* secara efektif dan tidak sekadar menebak kelas mayoritas.

Tingginya nilai bobot fitur jaccard_tag (β_2 = 4,352532) dibandingkan jaccard_genre (β_1 = 2,307600) mengindikasikan bahwa kesamaan tema cerita (*tag*) memberikan kontribusi yang lebih dominan dalam menentukan kemiripan komik daripada sekadar kategori cerita (genre). Performa model ini juga sejalan dengan penelitian rujukan oleh [21] yang memperoleh akurasi sebesar 85,5% menggunakan algoritma *Logistic Regression*. Hal ini mengonfirmasi bahwa pendekatan *Content-Based Filtering* mengombinasikan *Jaccard Similarity* dan *Logistic Regression* sangat layak diterapkan untuk klasifikasi kemiripan komik.

Berdasarkan hasil pengujian pada Tabel 6, seluruh skenario pengujian menunjukkan hasil yang sesuai dengan yang diharapkan (berhasil), sehingga dapat disimpulkan bahwa fitur rekomendasi komik pada aplikasi TeenToon telah berfungsi dengan baik sesuai rancangan sistem. Meskipun sistem dirancang untuk menghasilkan rekomendasi Top-5 pada aplikasi TeenToon, evaluasi yang dilakukan dalam penelitian ini berfokus pada performa klasifikasi. Evaluasi berbasis metrik rekomendasi seperti *Precision*@5, *Recall*@5, dan metrik lainnya belum dapat disajikan karena keterbatasan interaksi pengguna.

2. Kesimpulan

Berdasarkan hasil penelitian dan pengembangan yang dilakukan pada aplikasi TeenToon, pembentukan *dataset* dari 25 komik menghasilkan 300 pasangan komik yang divalidasi melalui *majority vote* oleh tiga validator. Berdasarkan *dataset* tersebut, proses implementasi *Content-Based Filtering* menggunakan *Jaccard Similarity* dan algoritma *Logistic Regression* terbukti mampu melakukan klasifikasi tingkat kemiripan komik dengan *accuracy* sebesar 88,46%, *precision* sebesar 71,43%, recall sebesar 100%, F1-score sebesar 83,33% dan nilai ROC-AUC yang mencapai 0.982. Selain itu, integrasi model pada *frontend* Android berjalan dengan baik, di mana antarmuka aplikasi secara otomatis memicu pemanggilan data saat pengguna membuka halaman detail komik dan menampilkan daftar komik serupa secara *descending* berdasarkan skor kemiripan tertinggi. Hasil pengujian sistem menggunakan metode *black box testing* terhadap empat skenario kombinasi tingkat kemiripan genre dan *tag* menunjukkan bahwa seluruh skenario pengujian berjalan sesuai dengan yang diharapkan, sehingga fitur rekomendasi komik pada aplikasi TeenToon dinyatakan telah berfungsi dengan baik.

Untuk penelitian selanjutnya, disarankan untuk memperluas atribut pembentuk kemiripan tidak hanya terbatas pada genre dan *tag*, tetapi juga memanfaatkan data interaksi implisit pengguna (seperti riwayat bacaan, durasi membaca komik, atau rating komik), pengembangan juga dapat dilakukan dengan menerapkan metode *Hybrid Recommendation System* atau mencoba algoritma klasifikasi lain (seperti Random Forest, dan sebagainya) guna membandingkan performa rekomendasi pada dataset skala besar. Selain itu, dalam pengembangan sistem rekomendasi disarankan melakukan pengujian sistem menggunakan metrik evaluasi sistem rekomendasi seperti *Precision*@5, *Recall*@5 dan sebagainya apabila data interaksi pengguna memadai.

Daftar Pustaka

[1] Sumardhiyati R. Syarif, Danar Aswim, and Abdullah Muis Kasim, “Pengaruh Penggunaan Handphone Terhadap Minat Baca Siswa Kelas VII di MTs. Muhammadiyah Wuring,” *JURNAL PENDIDIKAN DAN ILMU SOSIAL (JUPENDIS)*, vol. 2, no. 1, 2024, doi: 10.54066/jupendis.v2i1.1713.

[2] J. Purba, B. K. Putrawan, R. Setiawan, D. P. Tindoilo, and J. A. Tan, “Menumbuhkan Sikap Gemar Membaca Alkitab Pada Siswa Kelas 8 Sekolah Menengah Pertama BOPKRI 3 Yogyakarta,” *SERVIRE: Jurnal Pengabdian Kepada Masyarakat*, vol. 4, no. 1, pp. 33–48, 2024, doi: 10.46362/servire.v4i1.203.

[3] F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook: Third Edition*, 3rd ed. Springer Cham, 2022. doi: 10.1007/978-1-0716-2197-4.

[4] D. Roy and M. Dutta, “A systematic review and research perspective on recommender systems,” *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00592-5.

[5] Roni Putra, Sumema, and Z. Mustika, “Similarity Judul Tugas Akhir Menggunakan Metoda Cosine, Jaccard, Rabin-Karp Pada Jurusan TI,” *JITSI : Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 5, no. 4, 2024, doi: 10.62527/jitsi.6.1.291.

[6] I. M. Siregar, D. Pratama, and C. Himawan, “Penggunaan Jaccard Similarity Coefficient dalam Optimasi Proses Rekrutmen Karyawan Berbasis Profil dan Kompetensi,” *SINTECH (Science and Information Technology) Journal*, vol. 7, no. 2, 2024, doi: 10.31598/sintechjournal.v7i2.1617.

[7] Santi Purwaningrum, Agus Susanto, and Ari Kristiningsih, “Pengaruh Synonym Recognition dalam Deteksi Kemiripan Teks Menggunakan Winnowing dan Cosine Similarity,” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 12, no. 3, 2023, doi: 10.22146/jnteti.v12i3.6375.

[8] H. Sucipto, A. D. Indriyanti, K. Mauludi, A. H. Mujianto, M. F. Rizal, and C. Mashuri, “Sistem Pengukuran Tingkat Kemiripan Judul dan Abstrak Skripsi Menggunakan Algoritma Winnowing dan Dice Similarity,” *Journal of Information System Research (JOSH)*, vol. 6, no. 1, 2024, doi: 10.47065/josh.v6i1.5948.

[9] R. Alana and A. Hartanto, “Implementation of Content Based Filtering Algorithm in Comic Recommendation System,” *SISTEMASI*, vol. 13, no. 4, pp. 1344–1355, 2024, doi: 10.32520/stmsi.v13i4.2944.

[10] W. F. A. M. Irwan, B. Santoso, and S. Kacung, “IMPLEMENTASI METODE CONTENT BASED FILTERING MENGGUNAKAN SYNOPSIS SIMILARITY UNTUK PEMILIHAN ANIME,” *Jurnal Informatika Teknologi dan Sains (Jinteks)*, vol. 7, no. 2, pp. 947–955, 2025, doi: 10.51401/jinteks.v7i2.5815.

[11] D. Fahrizal and A. H. Hasugian, “INSOLOGI: Jurnal Sains dan Teknologi Sistem Rekomendasi TV Series Berdasarkan Genre Menggunakan Algoritma KNN,” *Media Cetak*, vol. 4, no. 4, pp. 895–906, 2025, doi: 10.55123/insologi.v4i4.6225.

[12] D. Triady, Izmy Alwiah Musdar, and Hendra Surasa, “PENGUJIAN BLACKBOX PADA WEBSITE WORKER’S MENGGUNAKAN METODE EQUIVALENCE PARTITIONING,” *KHARISMA Tech*, vol. 18, no. 1, 2023, doi: 10.55645/kharismatech.v18i1.292.

[13] O. Levin, *Discrete Mathematics: An Open Introduction, Fourth Edition*. 2025. doi: 10.1201/9781003589907.

[14] A. Vanacore and M. S. Pellegrino, “Benchmarking procedures for characterizing the extent of rater agreement: a comparative study,” *Qual. Reliab. Eng. Int.*, vol. 38, no. 3, 2022, doi: 10.1002/qre.2982.

[15] W. Winarto, I. Alwiah Musdar, and H. Hasniati, “ANALISIS SENTIMEN BAKAL CALON PRESIDEN 2024 MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE PADA TWITTER,” *KHARISMA Tech*, vol. 19, no. 1, 2023, doi: 10.55645/kharismatech.v19i1.464.

[16] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets*, 3nd ed. Cambridge University Press, 2020. doi: 10.1017/9781108684163.

[17] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R Second Edition*, 2nd ed. Springer, 2021. doi: 10.1007/978-1-0716-1418-1.

[18] M. R. Maulana, E. B. Susanto, and P. A. Christianto, “Pengujian Black Box dengan Teknik Equivalence Partitioning pada Aplikasi Monitoring Pemberian Obat Filariasis Berbasis Android,” *Media Online*, vol. 4, no. 4, 2024, doi: 10.30865/klik.v4i4.1603.

[19] M. S. S. Thayf, H. Angriani, Marlina, and Arianti, *Buku Ajar Pengantar Sains Data*, 1st ed. Media Sains Indonesia, 2025.

[20] Sutarman, R. Siringoringo, D. Arisandi, E. Kurniawan, and E. B. Nababan, “Model Klasifikasi Dengan Logistic Regression Dan Recursive Feature Elimination Pada Data Tidak Seimbang,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 4, 2024, doi: 10.25126/jtiik.1148198.

[21] R. Syahranita, S. Suhartono, and S. Zaman, “Prediksi Kategori Kelulusan Mahasiswa Menggunakan Metode Regresi Logistik Multinomial,” *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 8, no. 2, pp. 102–111, 2023, doi: 10.14421/jiska.2023.8.2.102-111.