

# Sistem Rekomendasi Manga (Komik Jepang) Menggunakan Metode Content-based Filtering

Maharani Putri Suari<sup>a1</sup> dan I Ketut Gede Suhartana<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Universitas Udayana  
Jimbaran, Badung, Bali, Indonesia  
<sup>1</sup>maharani802@gmail.com  
<sup>2</sup>ikg.suhartana@unud.ac.id

## Abstract

*Japanese comics or Manga are works of art from Japan that have spread almost all over the world. This manga is generally favored by children and teenagers because of its interesting art styles and stories. Due to advances in science and technology, manga can already be accessed on digital platforms in softcopy form. This certainly makes it easier to promote manga as well as read manga. Therefore, the manga recommendation system based on genre is designed to help readers get recommendations for other mangas according to their manga preferences using genre and content-based filtering methods with a dataset source from Kaggle website which has a dataset that contains 935 manga. The precision value obtained is 0.6 or 60%.*

**Keywords:** Recommendation System, Content-based Filtering, TF-IDF, Cosine, Precision, Manga

## 1. Pendahuluan

Komik Jepang merupakan salah satu budaya yang berasal dari Negara Jepang yang banyak digemari oleh masyarakat Indonesia terutama anak-anak hingga remaja yang mudah ditemui baik secara fisik yang biasanya terdapat di toko buku atau secara digital yang dapat diakses secara daring (online), dengan ciri khas berupa teknis penggambaran tokoh yang biasanya memiliki mata bulat lebar serta ukuran hidung dan mulut yang kecil, penggambaran latar belakang yang penuh dengan gambar dan tulisan dengan tujuan untuk menekankan situasi di dalam cerita, serta bentuk dan urutan panel cerita yang cenderung dinamis [1]. *Manga* (漫画) (dibaca sebagai: man-ga atau ma-ng-ga) adalah sebutan untuk komik dalam Bahasa Jepang, sedangkan untuk yang menggambar manga disebut sebagai Mangaka (漫画家) (dibaca sebagai: man-ga-ka atau ma-ng-ga-ka), sehingga di luar Negara Jepang kata Manga ini digunakan secara khusus untuk menyatakan bahwa komik yang sedang dibahas berasal dari Negara Jepang [2]. Osamu Tezuka (1928-1989) adalah seseorang yang sangat berpengaruh di dunia Manga dengan karyanya yang terkenal, yaitu "Astro Boy" dan manga adaptasi dari novel "Treasure Island" ciptaan Robert Louis Stevenson. Tezuka disebut sebagai "Bapak Manga" karena ia telah menciptakan manga yang sangat banyak (700 manga dengan kurang lebih 170.000 halaman) dan ia juga dianggap sebagai pionir teknik dan genre dalam Manga. Pada tahun 1960-an, majalah komik untuk remaja muncul seperti contohnya Manga Action (1967) dan Young Comic (1967). Majalah-majalah komik tersebut pada umumnya memiliki tebal sekitar 200-850 halaman. Kemudian cerita-cerita dari majalah tersebut disatukan dan dicetak menjadi buku berukuran biasa, yang bernama "tankoubun" yang artinya volume. *Tankoubun* ini yang akhirnya diterjemahkan ke dalam bahasa-bahasa negara lain, termasuk Negara Indonesia. Kemudian dari *tankoubun*, Manga juga ikut diterjemahkan, dan manga yang sangat populer yang biasanya diadaptasi menjadi Anime (animasi kartun buatan Jepang) [2]. Berkat Anime inilah penyebaran Manga ke luar negeri juga semakin meningkat, apalagi di era saat ini yang sudah menggunakan internet, penyebaran dan pengaksesan manga dapat dilakukan dengan sangat mudah tanpa perlu mencari bentuk fisik dari manga yang diinginkan di toko buku.

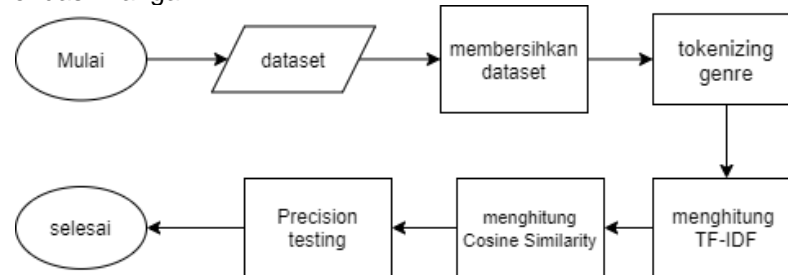
Karena popularitas manga yang mendunia ini, membuat sebagian orang tertarik untuk membaca manga, ditambah akses manga yang sudah banyak secara daring pun mempermudah pembaca untuk mencari manga yang serupa dengan manga yang mereka sukai berdasarkan genre. Oleh karena itu, alasan dilakukannya pembuatan sistem rekomendasi *manga* menggunakan *content-based filtering*

adalah untuk membantu memberikan rekomendasi manga, sehingga pembaca dapat dengan mudah memilih manga apa yang akan dipilih berdasarkan genre dari manga yang mereka sukai. Penelitian dari Muhammad Alkaff dengan judul “Sistem Rekomendasi Buku Menggunakan Weighted Tree Similarity dan Content Based Filtering” menjadi referensi untuk penelitian ini, dimana penelitian ini memiliki nilai *precision* sebesar 88%.

## 2. Metode Penelitian

### 2.1 Rancangan Penelitian

Sebelum memulai penelitian, dibuat rancangan penelitian sebagai alur untuk memudahkan dalam melakukan penelitian penggunaan metode *Content-based filtering* untuk membuat sistem rekomendasi manga.



**Gambar 1.** Alur rancangan penelitian

### 2.2 Pengambilan dan Pembersihan Dataset

Dataset yang digunakan berasal dari *website* Kaggle, dimana dataset ini berformat file CSV. Dataset ini memiliki 1000 data manga yang bersumber dari situs MyAnimeList. Setelah dataset diunduh, dataset dicek dan dibersihkan seperti menghapus komik-komik yang tidak termasuk manga dan manga yang judulnya duplikat, maka jumlah akhir dari dataset manga yang digunakan adalah 935 manga.

### 2.3 Tokenisasi

Tokenisasi bertujuan untuk memecah kalimat menjadi kata-kata terpisah, termasuk juga untuk menghilangkan simbol-simbol kata di dalam kalimat sehingga menjadi token. Bagian yang dilakukan tokenisasi di dataset ini adalah pada genre manga.

### 2.4 TF-IDF

Metode TF-IDF adalah metode pemberian bobot untuk hubungan sebuah kata dengan dokumen. Jika TF berfungsi untuk mengukur frekuensi kata yang ada di dalam sebuah dokumen dengan tetap memerhatikan panjangnya dokumen, maka IDF yang merupakan kebalikan dari DF (Document Frequency) berfungsi untuk menunjukkan kedekatan suatu token (kata) dengan dokumen. IDF akan bernilai rendah jika sebuah kata sering muncul di dalam dokumen, contohnya seperti kata hubung [3].

Rumus TF:

$$tf = 0,5 + 0,5 \times \frac{tf}{\max(tf)} \quad (1)$$

Rumus IDF:

$$idf_t = \log \left( \frac{D}{df_t} \right) \quad (2)$$

Yang kemudian menjadi:

$$W_{a,t} = tf_{a,t} \times IDF_{a,t} \quad (3)$$

### 2.5 Cosine Similarity

*Cosine similarity* adalah metode yang digunakan untuk mengetahui nilai kedekatan dokumen teks, dimana perhitungan dimulai dengan menghitung dot product. Dot product adalah perhitungan sederhana untuk tiap komponen yang berasal dari kedua vektor [4].

Rumus *cosine similarity*:

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (4)$$

## 2.6 Content-based Filtering

Sistem rekomendasi *content-based filtering* ini memberikan rekomendasi suatu hal untuk pengguna yang berdasarkan pada hal-hal atau item yang disukai oleh pengguna dan atribut-atribut yang ada di item tersebut [5]. Item yang akan direkomendasikan akan memiliki aspek-aspek yang serupa dengan aspek yang terkandung di dalam item yang disukai pengguna, yang dimana hal ini dapat dilakukan dengan cara menghitung *similarity* antara kedua item tersebut.

## 2.7 Precision Testing

*Precision testing* digunakan untuk mengetahui seberapa besar ketepatan rekomendasi yang diberikan dengan input yang diberikan oleh pengguna. Oleh karena itu, *precision testing* ini penting untuk mengetahui apakah sistem rekomendasi yang telah dibuat mampu untuk memberikan rekomendasi yang baik untuk pengguna.

Persamaan dari precision:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

Keterangan:

TP = *true positive* (jumlah output yang sesuai / relevan)

FP = *false negative* (sisa dari keseluruhan output)

## 3. Hasil dan Pembahasan

### 3.1 Tokenisasi Genre Manga

Tujuan dari melakukan tokenisasi genre manga adalah memecah genre-genre yang ada di dalam sebuah manga sehingga menjadi token yang berguna dalam mengekstrak makna dari genre tersebut, sehingga akan mudah untuk dilakukan pencarian kesamaan satu manga dengan manga lainnya berdasarkan genre.

```
"kids", "kids", "magic", "magic", "martial", "mecha",  
"military", "military", "music", "mystery", "parody",  
"police", "police", "psychological", "psychological",  
"romance", "romance", "samurai", "samurai", "school",  
"school", "sci-fi", "sci-fi", "seinen", "seinen",  
"shoujo", "shoujo", "shoujo", "shounen", "shounen",  
"shounen", "slice", "space", "space", "sports",  
"sports", "super", "supernatural", "supernatural",  
"thriller", "vampire", "vampire", "action", "adventure",
```

Gambar 2. Sample hasil tokenisasi genre manga

### 3.2 Penghitungan TF-IDF

Penghitungan TF-IDF ini dilakukan dengan tujuan memberikan bobot nilai antara judul manga dengan genre-genre yang terdapat pada manga. Penghitungan TF-IDF ini menggunakan *library*

sckit-learn dengan modulnya yaitu TfidfVectorizer yang tersedia di Bahasa Pemrograman Python.

### 3.3 Penghitungan *Cosine Similarity*

Setelah menghitung TF-IDF, dilakukan kalkulasi *cosine similarity* menggunakan *library* sckit-learn dengan modulnya *cosine\_similarity* yang juga tersedia di Bahasa Pemrograman Python. Ini adalah hasil penghitungan *cosine similarity* menggunakan matriks TF-IDF.

```
array([[1.         , 0.46252832, 0.2200812 , ..., 0.07099688, 0.         ,
        ],
       [0.46252832, 1.         , 0.28663895, ..., 0.13190594, 0.         ,
        ],
       [0.2200812 , 0.28663895, 1.         , ..., 0.12994014, 0.         ,
        ],
       ...,
       [0.07099688, 0.13190594, 0.12994014, ..., 1.         , 0.14520946,
        ],
       [0.         , 0.         , 0.         , ..., 0.14520946, 1.         ,
        ],
       [0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        ],
       [1.         , 0.         , 0.         , ..., 0.         , 0.         ]])
```

Gambar 3. Hasil penghitungan *cosine similarity*

Tujuan dari penghitungan *cosine similarity* ini adalah untuk mengetahui nilai kesamaan satu item dengan item lainnya di dalam dataset, dimana konteksnya adalah komik jepang atau manga sebagai itemnya. Berikut ini adalah sample manga yang dibandingkan dengan manga lain untuk melihat nilai kesamaannya. Semakin menuju angka 1 nilainya, maka kedua manga tersebut akan semakin mirip. Manga yang mirip ini yang akan dijadikan manga rekomendasi kepada pengguna.

Title	Ookami Kodomo no Ame to Yuki	Kimi to Natsu no Naka	Kiss yori mo Hayaku	Tobaku Datenroku Kaiji	Blue Giant
Fullmetal Alchemist	0.000000	0.000000	0.116196	0.000000	0.000000
Aishiteruze Baby	0.122782	0.118305	1.000000	0.000000	0.000000
Akagami no Shirayuki-hime	0.467198	0.136555	0.520295	0.168651	0.000000
Saiyuuiki Gaiden	0.000000	0.000000	0.000000	0.000000	0.000000
Boku to Kimi no Taisetsu na Hanashi	0.000000	0.113845	0.493734	0.000000	0.000000
Gunnm	0.184610	0.000000	0.000000	0.192305	0.180293
Toaru Majutsu no Index Gaiden: Toaru Kagaku no Railgun	0.000000	0.000000	0.000000	0.000000	0.000000
Holyland	0.105177	0.000000	0.000000	0.352935	0.102718
Ansatsu Kyoushitsu	0.000000	0.120395	0.179363	0.000000	0.000000
Hirunaka no Ryuusei	0.381484	0.500811	0.395874	0.000000	0.000000

Gambar 4. Tabel nilai *cosine similarity* pada manga

Pada tabel tersebut manga berjudul “Aishiteruze Baby” dengan manga berjudul “Kiss yori mo Hayaku” memiliki nilai 1, dimana kedua ini memiliki kemiripan sangat baik dalam genre.

Pembuktiannya adalah melakukan input manga “Aishiteruze Baby” dan muncul manga berjudul “Kiss yori mo Hayaku”.

```

▶ manga = (input('Input judul manga: '))
  recommendations(manga)

Input judul manga: Aishiteruze Baby
['Moe Kare!!',
 'I'm no Angel!',
 'Special A',
 'Aishiteruze Baby',
 'Lovely Complex',
 'Kiss yori mo Hayaku',
 'Kareishi Kanojo no Jijou',
 'Kaichou wa Maid-sama!',
 'Nekota no Koto ga Kininatte Shikatanai.',
 'Tonari no Kaibutsu-kun']
    
```

**Gambar 5.** Pembuktian nilai *cosine similarity*

### 3.4 Precision testing

Pada tahap ini akan dilakukan penghitungan presisi metode untuk rekomendasi manga dengan menggunakan sebuah manga input yang akan dibandingkan dengan 10 manga rekomendasi dari output. Manga “Detective Conan” akan dijadikan sebagai input.

```

Input judul manga: Detective Conan
['City Hunter',
 'Shibatora',
 'Pandora Hearts',
 'Karakuri Circus',
 'Binbougami ga!',
 'Konosuba: God's Blessing on This Wonderful World!',
 'Kieli: Shisha-tachi wa Kouya ni Nemuru',
 'Koten-bu Series',
 'Senyuu.',
 'Rave']
    
```

**Gambar 6.** Precision testing

Dengan menggunakan persamaan (5) dan nilai batas relevan yang digunakan adalah  $\geq 0.6$ , maka hasil yang diperoleh adalah:

Judul Manga	Nilai Precision
City Hunter	1.0
Shibatora	0.5
Pandora Hearts	0.6
Karakuri Circus	1.0
Binbougami ga!	0.5
Konosuba: God's Blessing on This Wonderful World!	0.5
Kieli: Shisha-tachi wa Kouya ni Nemuru	0.25
Koten-bu Series	0.6
Senyuu.	0.75
Rave	0.75

**Tabel 1.** Hasil nilai precision

Dari tabel tersebut, nilai *precision* secara keseluruhan adalah:

$$\frac{6}{6+4} = \frac{6}{10} = 0.6 \text{ atau akurasi sebesar } 60\%$$

#### 4. Kesimpulan

Kesimpulan yang diperoleh berdasarkan hasil penelitian yang diperoleh dari menggunakan metode *content-based filtering* pada dataset komik Jepang (manga) dengan *precision* sebagai penghitung keakuratannya adalah:

- 4.1 Metode *content-based filtering* yang digabungkan dengan TF-IDF dan *cosine similarity* dapat menjadi metode sistem rekomendasi yang baik bagi pengguna karena item yang direkomendasikan sebagian besar serupa dengan item yang menjadi *input* yaitu dari segi genrenya.
- 4.2 Penggunaan *precision* sebagai pengukur keakuratan terbilang cukup layak karena dapat menghitung keakuratan dari sistem rekomendasi menggunakan hasil yang diberikan oleh sistem rekomendasi tersebut.
- 4.3 Dalam penelitian ini, untuk menghasilkan nilai *precision* yang lebih tinggi, dataset yang digunakan harus dalam jumlah besar, karena semakin banyak data manga yang ada, maka semakin besar peluang genre manga yang sama antara satu dengan yang lainnya, sehingga semakin banyak manga yang dapat direkomendasikan.
- 4.4 Saran untuk penelitian serupa yaitu pembuatan sistem rekomendasi dengan metode *content-based filtering* adalah sebaiknya menggunakan dataset dalam jumlah besar dan juga tentukan dengan tepat unsur dari item yang akan direkomendasikan dimana unsur tersebut dapat dijadikan sebagai perbandingan untuk melihat kemiripan satu item dengan item lainnya.

#### References

- [1] A. Saifudin, "Penggunaan Manga Humor dalam Pembelajaran Bahasa dan Penelitian Bahasa Jepang", JAPANEDU, vol. 2, no. 2, p. 99-113, 2017.
- [2] F. E. Widigdo, *Penyebaran Budaya Pop Jepang Melalui Anime dan Manga di Indonesia*, Universitas Pembangunan Nasional "Veteran" Yogyakarta, 2021.
- [3] P. N. Raharjo, A. Handojo, H. Juwiantho, "Sistem Rekomendasi Content Based Filtering Pekerjaan dan Tenaga Kerja Potensial menggunakan Cosine Similarity", Jurnal Infra, vol. 10, No. 2, 2022.
- [4] W. G. S. Parwita, "Pengujian Akurasi Sistem Rekomendasi Berbasis Content-Based Filtering", Jurnal Informatika Mulawarman, Vol. 14, No. 1, p. 27-32, 2019.
- [5] M. R. A. Zayyad, *Sistem Rekomendasi Buku Menggunakan Metode Content Based Filtering*, Universitas Islam Indonesia, 2021.