

Website Rekomendasi Anime Dengan Menggunakan Pendekatan Content-Based Filtering Berdasarkan Sinopsis

Arfal Razyah Suhendra^{a1}, I Gusti Ngurah Anom Cahyadi Putra, ST., M.Cs^{a2}

^aInformatika, Universitas Udayana
Jimbaran, Bali
¹razyaarf@gmail.com
²anom.cp@unud.ac.id

Abstract

Today there are various kinds of Anime titles that can be found to be able to provide appropriate recommendations to Anime fans. A recommendation system is designed that can provide recommendations for Anime titles based on the similarity of the synopsis that has been read by users. The design of this Recommendation System uses a Content Based Filtering approach using the TFIDF method to obtain features that represent each synopsis which then produces a TFIDF matrix. The TFIDF matrix is then used as input to calculate the degree of similarity between synopsis. The results of the calculation of the degree of similarity become the basis for the formation of a function that can provide recommendations for Anime titles to users.

Keywords: Recommendation system, Content-Based Filtering, Cosine Similarity, TF-IDF, Website

1. Introduction

Saat ini terdapat berbagai macam hiburan yang dapat diakses melalui internet. Salah satu hiburan yang tersedia adalah animasi jepang atau yang biasa disebut Anime. Anime merupakan animasi khas Jepang yang memiliki ciri pada gambar yang berwarna warni yang menampilkan berbagai macam tokoh serta lokasi dari cerita yang ditunjukkan, Anime dipengaruhi oleh gaya gambar manga komik khas Jepang. Saat ini terdapat berbagai macam judul Anime dari berbagai macam genre yang tersedia dan dapat ditonton dari layanan streaming berbayar. Hal tersebut menyebabkan penggemar Anime sukar untuk dapat menentukan judul yang sesuai dengan preferensi yang disukai.

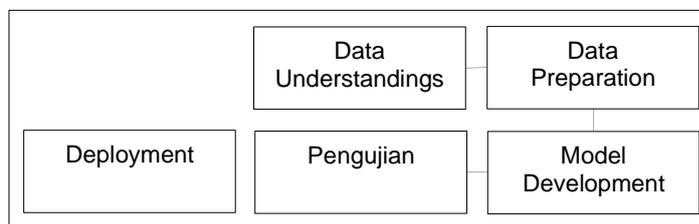
Oleh karena itu untuk dapat memberikan Anime yang sesuai dengan kegemaran dari pengguna dalam penelitian ini akan mengembangkan sistem rekomendasi yang dapat memberikan judul Anime yang sesuai kepada pengguna berdasarkan kesamaan sinopsis. Sistem rekomendasi dapat di definisikan sebagai program yang mencoba memberikan rekomendasi item yang paling cocok untuk pengguna tertentu dengan memprediksi minat pengguna pada item berdasarkan informasi terkait tentang item, pengguna dan interaksi antara item dan pengguna[6].

Sebagai salah satu aplikasi penyaringan informasi yang paling sukses, sistem rekomendasi telah menjadi cara yang efisien untuk memecahkan masalah kelebihan informasi. Tujuan dari sistem rekomendasi adalah untuk secara otomatis menghasilkan item yang disarankan untuk pengguna sesuai dengan preferensi historis mereka dan menghemat waktu pencarian mereka secara online dengan mengekstraksi data yang berguna.

Penelitian ini akan menggunakan dataset Anime-recommendation-database-2020 yang diperoleh dari website kaggle.com. Sinopsis yang diperoleh dari dataset tersebut akan diolah menjadi matriks TFIDF yang kemudian akan dihitung kesamaan antar sinopsis dengan metode *Cosine Similarity*. Hasil perhitungan kesamaan tersebut akan menghasilkan matriks kesamaan yang menunjukkan kesamaan antar judul. Setelah itu matrix akan di-*deploy* ke *web server* agar dapat memberikan rekomendasi ke pengguna.

2. Research Methods

Penelitian ini diawali dengan pengumpulan data dan dilanjutkan dengan penerapan dari metode yang digunakan.



Gambar 1 Alur pengembangan sistem

1. Data Understanding

Tahap ini dilakukan untuk memeriksa data, sehingga dapat mengidentifikasi masalah dalam data. Tahap ini memberikan pondasi analitik untuk sebuah penelitian dengan membuat ringkasan (*summary*) dan mengidentifikasi potensi masalah dalam data.

2. Data Preparation

Tahap Data Preparation secara garis besar adalah untuk memperbaiki masalah dalam data, kemudian membuat *variabel derived*. Kegiatan yang dilakukan antara lain: memilih kasus dan parameter yang akan dianalisis (*Select Data*), melakukan transformasi terhadap parameter tertentu (*Transformation*), dan melakukan pembersihan data agar data siap untuk tahap *modeling* (*Cleaning*).

3. Pembentukan Model

Secara garis besar untuk membuat model prediktif atau deskriptif. Pada tahap ini dilakukan metode statistika dan *Machine Learning* untuk penentuan terhadap teknik *data mining*, alat bantu *data mining*, dan algoritma *data mining* yang akan diterapkan. Lalu selanjutnya adalah melakukan penerapan teknik dan algoritma data mining tersebut kepada data dengan bantuan alat bantu.

4. Pengujian

Melakukan interpretasi terhadap hasil dari data mining yang dihasilkan dalam proses pemodelan pada tahap sebelumnya. Evaluasi dilakukan terhadap model yang diterapkan pada tahap sebelumnya dengan tujuan agar model yang ditentukan dapat sesuai dengan tujuan yang ingin dicapai dalam tahap pertama.

5. Deployment

Pada tahap ini model yang dihasilkan di-deploy ke web server.

2.1. Metode Pengumpulan data

Data yang digunakan pada penelitian ini merupakan data sekunder yang bersumber dari website kaggle.com. Dataset yang digunakan yang digunakan merupakan data yang disimpan pada file csv yang menyimpan MAL_ID yang merupakan id dari judul Anime, Name yang merupakan judul Anime, Score, genre dan synopsis.

2.2. Content Based Filtering

Content Based Filtering merupakan salah satu pendekatan yang digunakan dalam pengembangan sistem rekomendasi. Pada sistem rekomendasi *Content Based Filtering* bergantung pada pilihan sebelumnya pengguna. Content-based Recommendation menggunakan ketersediaan konten sebuah item sebagai basis dalam pemberian rekomendasi[2].

2.3. Cosine Similarity

Kemiripan antar dokumen dihitung menggunakan suatu fungsi ukuran kemiripan (*similarity measure*). *Cosine Similarity* digunakan untuk melakukan perhitungan kesamaan dari dokumen. Rumus yang digunakan oleh *cosine similarity* adalah:

$$\cos\alpha = \frac{A \cdot B}{|A||B|} \quad (1)$$

Keterangan :

A = Vektor A, yang akan dibandingkan kemiripannya

B = Vektor B, yang akan dibandingkan kemiripannya

$A \cdot B$ = *dot product* antara vektor A dan vektor B

|A| = panjang vektor A

|B| = panjang vektor B

|A||B| = *cross product* antara |A| dan |B|

2.4. TF IDF

TF-IDF merupakan kependekan dari *Term Frequency — Inverse Document Frequency*. TF-IDF biasa digunakan untuk mengubah data teks menjadi matriks fitur TF-IDF. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut[3]. Jumlah kemunculan sebuah kata dalam dokumen tertentu menunjukkan betapa pentingnya kata itu dalam dokumen itu. Frekuensi dokumen yang berisi kata tersebut menunjukkan seberapa umum kata tersebut. Jadi hubungan antara sebuah kata dan sebuah dokumen akan tinggi jika frekuensi kata dokumen tinggi dalam dokumen dan frekuensi seluruh dokumen yang mengandung kata di set pada dokumen.

3. Result and Discussion

3.1. Data Understandings

Melihat informasi kolom pada dataset *Anime_synopsis* dengan fungsi `info()`. Hasil dari fungsi `info` memberikan informasi bahwa terdapat satu tipe data integer pada kolom `MAL_ID` dan tipe data `object` untuk kolom lainnya. Dataset tersebut berisi informasi mengenai “MAL_ID” yang merupakan id dari judul Anime yang disimpan, “Name” yang merupakan judul dari “Anime”, “Score”, “Genres” dan “synopsis”.

3.2. Data Preparation

Pada tahap ini dilakukan pengecekan terhadap missing value. pengecekan untuk missing value menggunakan fungsi `isnull()`, kemudian menjumlahkan nilai pada kolom yang mengandung missing value. Dari Hasil pengecekan pada datasets *Anime_synopsis* terdapat 8 missing value pada kolom `synopsis`. Kolom yang mengandung nilai dengan value `NaN` kemudian diganti pada baris yang berisi missing value pada kolom `synopsis` dengan nilai string kosong. Hal ini dilakukan agar setiap baris pada kolom `synopsis` nantinya dapat direpresentasikan dengan TF-IDF. Ini dilakukan dengan fungsi `fillna()` yang akan mengisi nilai yang kosong pada kolom `synopsis` dengan string kosong.

3.3. Pembentukan Sistem Rekomendasi

Pada tahap ini dengan data yang telah dibersihkan dilakukan tokenisasi dengan fungsi `TfidfVectorizer()` dari library `sklearn` dengan masukan kolom `synopsis` dari datasets *Anime_synopsis*. Tokenisasi dilakukan untuk menemukan representasi fitur penting dari setiap `synopsis`. Proses ini akan menghitung kata yang terdapat di setiap `synopsis` kemudian Tokenisasi ini kemudian akan menghasilkan matriks yang nantinya data dari `synopsis` yang telah dilakukan tokenisasi dapat dihitung derajat kesamaan antar `synopsis` menggunakan metode *Cosine Similarity*.

Selanjutnya untuk menghitung derajat kesamaan (*similarity degree*) antar sinopsis dilakukan dengan menggunakan metode *Cosine Similarity*. Untuk menggunakan metode tersebut dilakukan dengan menggunakan *library* yang telah tersedia pada Sklearn. Hasil dari perhitungan tersebut akan menghasilkan matriks yang menunjukkan kesamaan antar sinopsis. Matriks tersebut akan menampung nilai 0 sampai dengan 1. Nilai 0 menunjukkan kedua sinopsis yang dihitung sama sekali berbeda. Sedangkan nilai 1 menunjukkan sinopsis yang dihitung, itu akan benar-benar mirip.

Name	Legendz: Yomigaeru Ryuuou Densetsu	Kaibutsu Gakushou	Ajin OVA	Akuei to Gacchinpo The Movie: Chelsea no Gyakushuu/Akuei to Mahou no Hammer	High School DxD New: Oppai, Tsutsumimasu!
Taboo Tattoo	0.055020	0.024575	0.014084	0.018867	0.010927
Dr. Rin ni Kittemite!	0.015444	0.007443	0.010784	0.003866	0.038140
Kyomu Senshi Miroku	0.013988	0.012051	0.025907	0.001978	0.031013
Himouto! Umaru-chan R	0.048182	0.014902	0.018723	0.018688	0.011810
Hiragana Danshi	0.037912	0.019453	0.010375	0.003310	0.014651

Gambar 2 Hasil kesamaan setiap judul

Setelah mendapatkan matriks kesamaan antar sinopsis pembuatan model dilanjutkan dengan membuat fungsi *Anime_recommender*. Fungsi ini akan mengambil baris dari index judul Anime yang dimasukkan pengguna, kemudian mengurutkan nilai terbesar hingga terkecil dari matriks kesamaan. Kemudian memberikan hasil berupa N judul Anime dengan nilai terbesar yang bersesuaian dengan judul Anime tersebut.

3.4. Pengujian

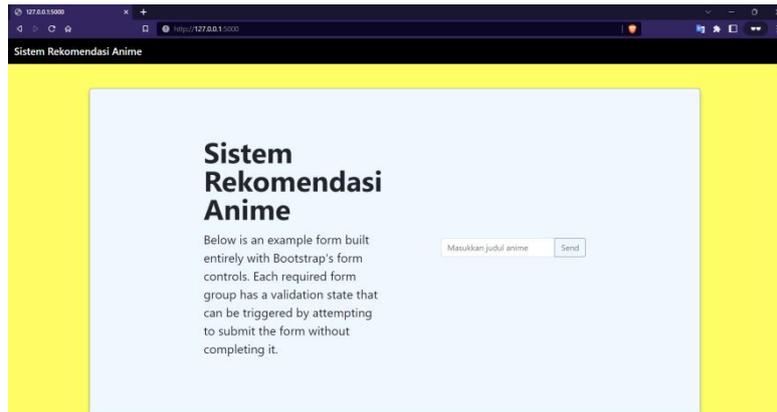
Pengujian dilakukan dengan memanggil fungsi *Anime_recommender* dengan memasukkan parameter berupa judul Anime yang menjadi acuan dalam memberikan rekomendasi. Judul anime yang digunakan untuk pengujian dan dimasukkan pada sistem rekomendasi adalah "One Punch Man". Hasil pengujian menunjukkan fungsi yang dihasilkan dapat memberikan judul yang mirip dengan judul Anime yang dimasukkan pada fungsi tersebut.

	Name	synopsis	Genres
0	One Punch Man: Road to Hero	Before Saitama became the man he is today, he ...	Action, Sci-Fi, Comedy, Parody, Super Power, S...
1	Aura: Maryuin Kouga Saigo no Tatakai	Ichirou Satou is an ordinary high school stude...	Supernatural, Drama, Romance, School
2	One Punch Man 2nd Season	In the wake of defeating Boros and his mighty ...	Action, Sci-Fi, Comedy, Parody, Super Power, S...
3	Yumeria	On his 16th birthday, Tomokazu Mikuri had a re...	Action, Comedy, Ecchi, Harem, Super Power
4	Dog Days'	Cinque returns to the land of Flonyard in Dog ...	Action, Adventure, Magic, Fantasy

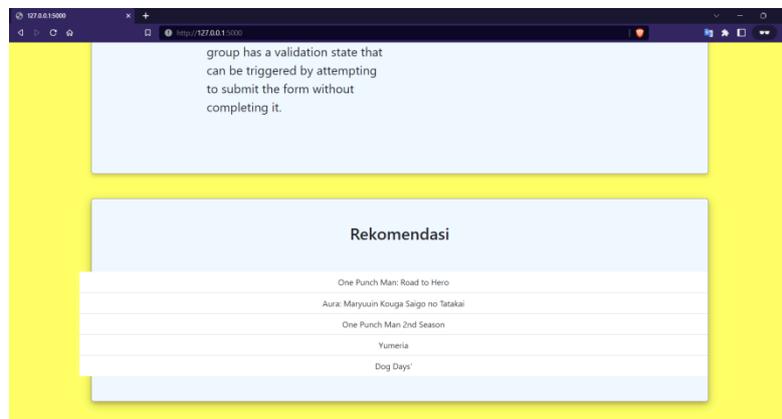
Gambar 3 Hasil pengujian fungsi Anime recommender

3.5. Deployment

Fungsi yang telah dihasilkan di *deploy* ke *web server*. Pada pengembangan web server menggunakan framework flask. Halaman utama dari website dapat dilihat pada gambar berikut.



Gambar 4 Tampilan Halaman utama website



Gambar 5 Tampilan hasil rekomendasi pada website

4. Conclusion

Pada penelitian ini telah dirancang sistem rekomendasi untuk rekomendasi judul Anime. Ini memungkinkan pengguna untuk mendapatkan judul Anime dengan sinopsis yang mirip dengan judul Anime yang telah dibaca oleh pengguna.

Penggunaan *cosine similarity* sebagai metode penghitung kesamaan dari sinopsis dapat memberikan matriks kesamaan yang menjadi dasar pemberian rekomendasi ke pengguna. Pengujian dilakukan dengan memasukkan judul Anime tertentu kemudian sistem yang dikembangkan dapat memberikan judul Anime berdasarkan judul yang telah diberikan tersebut.

References

- [1]F. Nugroho and M. Ismu Rahayu, "SISTEM REKOMENDASI PRODUK UKM DI KOTA BANDUNG MENGGUNAKAN ALGORITMA COLLABORATIVE FILTERING," *Jurnal Riset Sistem Informasi dan Teknologi Informasi (JURSISTEKNI)*, vol. 2, no. 3, pp. 23–31, Sep. 2020, doi: 10.52005/jursistekni.v2i3.63.
- [2]A. Wijaya and D. Alfian, "Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering Dan Content-Based Filtering," *Jurnal Computech & Bisnis (e-Journal)*, vol. 12, no. 1, pp. 11–27, Jun. 2018, doi: 10.55281/jcb.v12i1.167.
- [3]M. Nurjannah, H. Hamdani, and I. F. Astuti, "PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING,"

Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer, vol. 8, no. 3, pp. 110–113, Jun. 2016, doi: 10.30872/jim.v8i3.113.

[4]B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, “Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 1, pp. 306–312, Aug. 2017.

[5]Y. K. Ng, “CBRec: a book recommendation system for children using the matrix factorisation and content-based filtering approaches,” *International Journal of Business Intelligence and Data Mining*, vol. 16, no. 2, p. 129, 2020, doi: 10.1504/ijbidm.2020.104738.

[6]J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, “Recommender system application developments: A survey,” *Decision Support Systems*, vol. 74, pp. 12–32, Jun. 2015, doi: 10.1016/j.dss.2015.03.008.

[7]P. S. Adi, “Sistem Rekomendasi Nilai Mata Kuliah Menggunakan Metode Content-based Filtering,” *Seminar Nasional Informatika 2010 (semnasIF 2010)*, vol. 1, no. 1, May 2010.

[8]P. Nastiti, “Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan,” *Teknika*, vol. 8, no. 1, pp. 1–10, Jun. 2019, doi: 10.34148/teknika.v8i1.139.

[9]W. G. Suka Parwita, M. H. Prami Swari, and W. Welda, “PERANCANGAN SISTEM REKOMENDASI DOKUMEN DENGAN PENDEKATAN CONTENT-BASED FILTERING,” *Computer Engineering, Science and System Journal*, vol. 3, no. 1, p. 65, Jan. 2018, doi: 10.24114/cess.v3i1.7855.

[10]M. W. Putri, A. Muchayan, and M. Kamisutara, “Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF,” *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 5, no. 3, pp. 229–236, Sep. 2020, doi: 10.31328/jointecs.v5i3.1563.

[11]M. Fajriansyah, P. P. Adikara, and A. W. Widodo, “Sistem Rekomendasi Film Menggunakan Content Based Filtering,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 6, pp. 2188–2199, 2021.

[12]M. M. Wiputra and Y. J. Shandi, “Perancangan Sistem Rekomendasi Menggunakan Metode Collaborative Filtering dengan Studi Kasus Perancangan Website Rekomendasi Film,” *Media Informatika*, vol. 20, no. 1, pp. 1–18, Mar. 2021, doi: 10.37595/mediainfo.v20i1.54.

[13]N. Ula, C. Setianingsih, and R. A. Nugrahaeni, “Sistem Rekomendasi Lagu Dengan Metode Content Based Filtering Berbasis Website,” *eProceedings of Engineering*, vol. 8, no. 6, Dec. 2021, doi: 17229.

[14]R. A. Misnadin, “Sistem Rekomendasi Abstrak Dokumen dengan Content Based Filtering,” Jan. 01, 2017. <http://etd.repository.ugm.ac.id/penelitian/detail/107496>

[15]P. N. Raharjo, A. Handojo, and H. Juwiantho, "Sistem Rekomendasi Content Based Filtering Pekerjaan Tenaga Kerja Potensial menggunakan CosineSimilarity," *Jurnal Infra*, vol. 10, no. 2.

Halaman ini sengaja dibiarkan kosong