Prosiding Seminar Implementasi Teknologi Informasi dan Komunikasi
Vol. 2, No. 1, Februari 2023
ISSN: 2809-9834, DOI: 10.31284/p.semtik.2023-1.4153

Sistem Rekomendasi Film dengan Menggunakan Pendekatan *Collaborative Filtering* Berdasarkan *Class*

Ibadurrohman, Septiyawan Rosetya Wardhana, Rinci Kembang Hapsari*

Institut Teknologi Adhi Tama Surabaya

ABSTRACT

A recommendation system is capable of presenting information that users are interested in. In this study, a film recommendation system using a class-based collaborative filtering approach was implemented using Term Frequency – Inverse Document Frequency – Inverse Class Frequency, which features pre-processing through various stages, including tokenizing, filtering, stopword removal and stemming. So that the data obtained is structured, then weighting will be carried out by combining the frequency of occurrence in sentences, the number of documents in a word that appears and the number of classes in a comment that appears. Then, the concept results are multiplied, which will later be taken as the highest value due to film recommendations.

Keywords

Collaborative Filtering,
Film,
Python,
Tf-idf-icf
System Recommendation

ABSTRAK

Sistem rekomendasi adalah sistem yang mampu menyajikan informasi yang pengguna minati. Pada penelitian ini sistem rekomendasi film dengan menggunakan pendekatan collaborative filtering berdasarkan class diterapkan dengan penggunaan Term Frequency – Inverse Document Frequency – Inverse Class Frequency, yang mana fitur pre-processing melalui berbagai tahapan, diantaranya tokenizing, filtering, stopword removal serta stemming. Sehingga data yang didapat terstruktur, selanjutnya akan dilakukan pembobotan dengan menggabungkan antara frekuensi kemunculan pada kalimat, banyak dokumen pada suatu kata yang muncul dan banyaknya kelas pada suatu kata yang muncul. Kemudian dari hasil konsep tersebut dikalikan yang nantinya akan diambil nilai tertinggi sebagai hasil rekomendasi film.

PENDAHULUAN

Film merupakan salah satu karya seni yang dibuat dari suatu imajinasi dan kreatifitas orang-orang yang ikut serta dalam proses pembuatan film. Sebagai suatu karya seni, film terbukti mampu menjadi sarana untuk mewujudkan kemampuan kreatif yang sanggup untuk menciptakan suatu realitas rekaan sebagai bandingan terhadap realitas. Realitas imajiner dalam film mampu menyajikan bentuk keindahan ataupun hanya untuk hiburan semata bagi penikmatnya.

Film merupakan salah satu dunia hiburan yang digemari oleh masyarakat. Semakin ramainya penggemar film perlu diberikan informasi terkait film-film supaya penonton tertarik dn berminat menonton film yang ditawarkan. Genre film merupakan bentuk atau kategori dari beberapa film tertentu yang mempunyai kesamaan bentuk, tema, suasana dan lain sebagainya. Ada beberapa genre film, diantaranya : action, adventure, comedy, horror, romance, drama dan lainnya.

Banyaknya genre dan judul film yang telah rilis serta keberadaan informasi film yang tersedia seharusnya bermanfaat dalam memberikan kemudahan masyarakat, tetapi kenyataannya banyak pilihan membuat masyarakat bingung dalam menentukan pilihan film mana yang akan ditonton. Dalam mengatasi hal tersebut diperlukan sebuah sistem rekomendasi untuk penentuan pemilihan film yang akan ditonton.

Sistem rekomendasi film telah banyak diteliti, diantaranya sistem rekomendasi film menggunakan kombinasi dari bisecting K-Means dan *Collaborative Filtering*. Data film yang digunakan pada penelitian ini berasal dari MovieLens yang terdiri dari 100.000 rating dari 668 user untuk 10329 judul film dalam 18 genre film. Proses training terdiri dari proses cluster dengan algoritma bisecting K-Means dan perhitungan nilai similarity dengan collaborative filtering (*item-based dan user-based*) [1].

Sistem rekomendasi film berbasis website dengan metode *prototyping* menggunakan algoritma K-Nearest Neighbors. Dimana K-NN digunakan untuk mengklasifikasikan objek baru berdasarkan atribut dan sampel dari data training. Metode pembangunan sistem prototype memiliki

beberapa tahapan yaitu communication, quick plan and modelling design, construction of prototype, serta deployment delivery and feedback[2].

Sistem rekomendasi film menggunakan *content based filtering* dengan mencari kemiripan bobot dari term pada bag of words hasil preprocessing sinopsis film dan judul film. Pembobotan dilakukan menggunakan metode TF-IDF yang telah dinormalisasi. Kemudian hasil pembobotan akan melalui tahap *cosine similarity* untuk mencari kemiripan berdasarkan bobot dan diakhiri dengan filtering berdasarkan genre. Hasil pengujian yang dilakukan dengan melibatkan tiga partisipan dengan total jumlah film sebanyak 4000 judul film didapatkan nilai akurasi menggunakan *mean average precision* @K (MAP@K) sebesar 0.823254[3]. Berdasarkan dari penelitian sebelumnya maka dalam penelitian ini dmembangun system rekomendasi film dengan menggunakan pendekatan *collaborative filtering* berbasiskan *class.*

TINJAUAN PUSTAKA

Film

Menurut Kamus Besar Bahasa Indonesia, film memiliki dua pengertian. Pertama, film adalah selaput tipis yang terbuat dari seluloid yang digunakan sebagai tempat gambar negatif (yang nantinya akan dibuat potret) atau sebagai tempat gambar positif (yang nantinya diputar di bioskop). Kedua, film juga memiliki arti cerita dalam hidup. Serta tentang film ini tak hanya menjadi sarana atau media kreasi bagi para seniman, melain juga dapat menjadi profesi yang menguntungkan dalam hal industrinya.

Collaborative Filtering

Sistem rekomendasi dalam berbagai aplikasi telah mencoba memberikan rekomendasi yang akurat kepada pengguna untuk memenuhi kebutuhan pengguna dan memberikan manfaat yang lebih tinggi bagi perusahaan. *Collaborative filtering* adalah teknologi yang efektif dan terkenal dalam sistem rekomendasi. Ide utama dalam sistem rekomendasi *collaborative filtering* adalah untuk memanfaatkan riwayat opini pengguna aktif lain untuk memprediksi item yang mungkin akan disukai/diminati oleh seorang pengguna [4].

Collaborative filtering merupakan proses penyaringan atau pengevaluasian item menggunakan opini orang lain. Salah satu teknik dalam sistem rekomendasi yang digunakan untuk menyarankan suatu item dari user kepada user lain yang memiliki kemiripan selera[5]. Collaborative filtering (CF) adalah suatu metode dalam membuat prediksi otomatis untuk memperkirakan ketertarikan atau selera seorang user terhadap suatu item dengan cara mengumpulkan informasi dari user-user yang lain yang direpresentasikan dalam bentuk nilai rating.

Terdapat dua kategori dan satu gabungan yang terbagi dalam *Collaborative Filtering* yaitu memory based, model based dan hybrid recommendation system yang tujuannya mengatasi masalah yang nantinya terdapat pada kedua kategori tersebut.

Algoritma *TF/IDF (Term Frequency – Inverse Document Frequency)*

TF-IDF dipakai untuk menentukan suatu nilai frekuensi suatu kata pada sebuah artikel atau dokumen serta frekuensi dalam banyak dokumen. TF-IDF (Term Frequency-Inverse Document Frequency) adalah metode Statistic numeric yang mampu menggambarkan perlunya suatu kata pada suatu dokumen. Yang sering dipakai untuk faktor bobot pada penambangan teks dan pencarian informasi. Nilai pada TF-IDF berkembang secara proporsional dengan dasar banyaknya kata yang sering muncul pada suatu dokumen, akan tetapi harus bersamaan dengan frekuensi kata, pada korpus. Varietas dari skema pembobotan ini biasa diterapkan oleh mesin pencari untuk alat utama saat mencetak nilai (scoring) serta peringkat suatu relevansi dokumen yang diberikan user.

Prosedur pada penerapan TF-IDF ada suatu perbedaan kecil dalam seluruh aplikasinya, namun pendekatannya hampir sama. Nilai dari TF bisa diperoleh dengan persamaan[6] (1)

$$Idf_i = \log log \left(\frac{n_d}{df_i} \right) \quad (1)$$

Idf_i = *Inverse document frequency* untuk tiap term/kata; log = Operasi matematika yang merupakan kebalikan dari operasi pangkat; nd = Jumlah dokumen yang ada pada data training; dfi = Jumlah dokumen yang memiliki kata tersebut.

ICF (*Inverse Class Frequency*)

ICF lebih menitik beratkan pada munculnya term suatu kumpulan kategori/kelas, yang jarang muncul di banyak kelas yaitu term yang mempunyai nilai untuk klasifikasi. Setiap term mempunyai kepentingan yang diasumsikan memiliki proporsi yang sebaliknya dengan jumlah kelas yang mengandung term[7]. *Inverse Class Frequency (ICF)* adalah proses yang dilakukan saat pembobotan kata. Bertujuan untuk mencari bobot pada setiap kata. Perhitungan ICF diperoleh dengan persamaan (2).

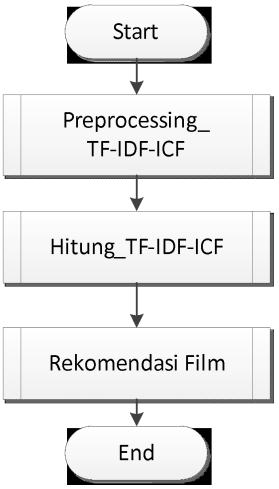
$$Icf_i = \log log \left(\frac{nc}{cf_i} \right) \quad (2)$$

METODE

Gambaran Umum

Penelitian ini memiliki tujuan supaya pengguna dapat lebih mudah mendapatkan rekomendasi film yang sesuai dengan Genre yang diinginkan. Pengambilan dataset dilakukan secara manual pada situs film imdb.com. Pengujian dilakukan dengan jumlah data yang diuji adalah 3 Genre dan 6 film.

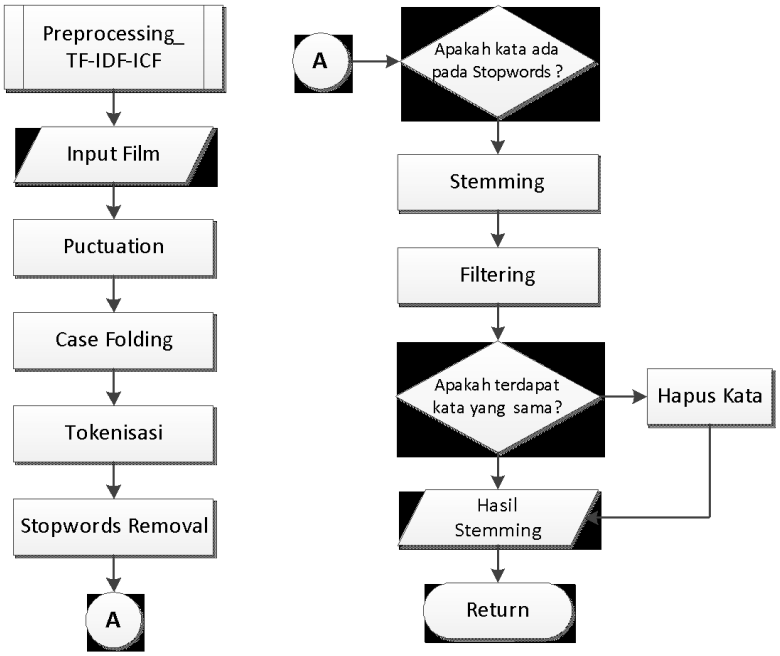
Gambaran umum alur penelitian ditunjukkan pada Gambar 1. Pada tahap pengujian user sudah menonton film avenger, dari film tersebut akan dijadikan history user. Pada history user akan diambil sinopsis film. Kemudian dilakukan perhitungan menggunakan TF-IDF-ICF pada synopsis yg terdapat pada history user dan database untuk memperoleh bobot kata yang akan digunakan untuk menentukan ranking film yang nantinya direkomendasikan kepada user.



Gambar 1. Alur utama sistem rekomendasi Film

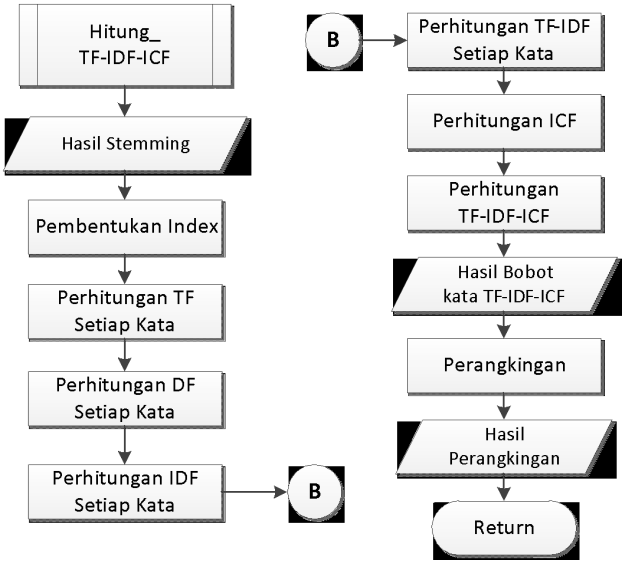
Tahapan p*reprocessing* dilakukan untuk menghindari data yang kurang sempurna, gangguan pada data, dan data-data yang tidak konsisten, yang ditunjukkan pada Gambar 2. Pertama dilakukan *Punctuation* yaitu proses membuang tanda baca dan angka pada inputan user. Kemudian melakukan *case folding* untuk mengubah seluruh kata menjadi huruf kecil. Proses tokenisasi dilakukan untuk memecah kalimat menjadi kata (*Term*). Kemudian dilakukan *Stopwords Removal* untuk membuang kata-kata yang tidak dibutuhkan dalam proses nantinya. Setelah proses diatas

dijalankan barulah dilakukan Stemming untuk merubah kata dasar (*Root Word*). Dan dilakukan Filtering untuk membuang kata-kata yang sama.



Gambar 2. Flowchart sub proses *Preprocessing*

TF-IDF-ICF (Term Frequency-Invers Dokumen Frequency-Invers Class Frequency) merupakan metode statistic numeric yang mencerminkan seberapa pentingnya sebuah kata dalam sebuah dokumen atau korpus. Alur sub proses perhitungan TF-IDF-ICF ditunjukkan pada Gambar 3.



Gambar 3. Flowchart sub proses Hitung TF-IDF-ICF

HASIL DAN PEMBAHASAN

Hitungan TF-IDF

TF-IDF-ICF (*Term Frequency-Invers Dokumen Frequency-Invers Class Frequency*) merupakan metode *statistic numeric* yang mencerminkan seberapa pentingnya sebuah kata dalam sebuah dokumen.

Tabel 1. Bobot Kata

Term	W = TF * IDF * ICF					
	Histori User	**D1**	**D2**	**D3**	**D4**	**D5**
Alli	0	0,070074	0	0,070074	0	0
Artifact	0	0,070074	0	0,070074	0	0
Aveng	0	0.140148	0	0,140148	0	0
Collect	0	0,070074	0	0,070074	0	0
Cosmic	0	0,070074	0	0,070074	0	0
New	0	0	0	0	0	0
One	0	0,017065	0,017065	0,175065	0,017065	0
World	0,078131	0,039066	0	0,039066	0	0,078131
Total	0,078131	0,476575	0,017065	0,476575	0,017065	0,078131

Perankingan

Pada penelitian ini, peneliti memberikan contoh hasil perankingan berdasarkan bobot kata pada film, ditunjukkan pada Tabel 2.

Tabel 2. Data Perankingan Film

Judul Film	Bobot Kata
Avenger infinity war	0,476575
Aquaman	0,017065
Venom	0,017065
Deadpool	0,078131

KESIMPULAN

Berdasarkan hasil pengujian dalam penelitian, maka didapatkan beberapa kesimpulan sebagai berikut:

1. Dari hasil uji coba yang dirancang untuk merekomendasikan film dengan metode *Term Frequency - Inverse Document Frequency - Inverse Class Frequency* yang diusulkan terbukti mampu merekomendasikan film.
2. Rekomendasi film dengan metode Term Frequency - Inverse Document Frequency - Inverse Class Frequency, user mendapatkan rekomendasi dari sistem film apa yang sesuai dengan yang user inginkan.

DAFTAR PUSTAKA

[1] A. Halim, H. Gohzali, D. Maria Panjaitan, and I. Maulana, "Sistem Rekomendasi Filmmenggunakan Bisecting K-Means dan Collaborative Filtering," *Citisee*, vol. 4567789, no. 061, pp. 37–41, 2017.

[2] A. R. Fitrianti, A. Rohmani, and W. Widjanarto, "Sistem Rekomendasi Film Berbasis Website Dengan Metode Prototype Menggunakan Algoritma K-Nearest Neighbors (KNN)," *JOINS (Journal Inf. Syst.*, vol. 5, no. 2, pp. 278–287, 2020, doi: 10.33633/joins.v5i2.4168.

[3] M. Fajriansyah, P. P. Adikara, and A. W. Widodo, "Sistem Rekomendasi Film

Menggunakan Metode Content Based Filtering," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 6, pp. 2188–2199, 2021, [Online]. Available: http://e-journal.uajy.ac.id/20600/

[4] D. Nugraha, T. W. Purboyo, and R. A. Nugrahaeni, "Sistem Rekomendasi Film Menggunakan Metode User Based Collaborative Filtering," in *e-Proceeding of Engineering*, 2021, vol. 8, no. 5, pp. 6765–6775. [Online]. Available: https://ejournal.upi.edu/index.php/JATIKOM

[5] P. Sen, G. M. Namata, M. Bilgic, L. Getoor, B. Gallagher, and T. Eliassi-Rad, "Collective classification in network data," *AI Mag.*, vol. 29, no. 3, pp. 93–106, 2008, doi: 10.1609/aimag.v29i3.2157.

[6] A. Hidayat, "Impementasi Metode Term Frequency and Inverse Document Frequency Dan Marginal Relevance Untuk Monitoring Diskusi Online," *J. Sains, Teknol. dan Ind.*, vol. 13, no. 2, pp. 151–159, 2016.

[7] P. Goswami and V. Kamath, "The DF-ICF Algorithm- Modified TF-IDF," *Int. J. Comput. Appl.*, vol. 93, no. 13, pp. 28–30, 2014, doi: 10.5120/16276-6036.