

Sistem Rekomendasi Film Menggunakan Metode Collaborative Filtering dan K-Nearest Neighbors

Film Recommendation System Using Collaborative Filtering Method and K-Nearest Neighbors

¹Eggy Ryana Agustian, ²Munir, ³Eddy Prasetyo Nugroho

Program Studi Ilmu Komputer Departemen Pendidikan Ilmu Komputer Universitas Pendidikan Indonesia
Bandung, Indonesia

¹eggyryana@gmail.com, ²munir@upi.edu,

³eddypn@upi.edu

Abstrak— Industri film berkembang demikian pesat. Banyak jenis film yang beredar dan siap untuk disaksikan. Oleh sebab itu banyak orang kebingungan akan film apa yang ingin disaksikan. Maka, dibutuhkan sistem untuk merekomendasikan film agar sesuai dengan yang calon penonton inginkan. Penelitian ini bertujuan untuk mengetahui prediksi rating film pada suatu pengguna menggunakan metode *Collaborative Filtering* (CF). Algoritma CF dapat memprediksi rating suatu film berdasarkan relasi antar user lainnya. Sistem berbasis website sehingga pengguna dapat langsung melihat hasil keluaran film yang direkomendasikan untuk disaksikan.

Kata Kunci: *sistem rekomendasi, film, collaborative filtering.*

Abstract— The film industry is growing so rapidly. Many types of films are circulating and ready to be watched. Therefore, many people are confused about what film to watch. So, a system is needed to recommend films to suit what the potential audience wants. This study aims to determine the prediction of a user's film rating using the Collaborative Filtering (CF) method. The CF algorithm can predict the rating of a film based on the relationship between other users. The system is website-based so that users can immediately see the recommended movie to watch.

Keywords: *recommendation system, film, collaborative filtering.*

I. PENDAHULUAN

Industri film saat ini sudah menjadi industri yang terbilang besar. terdapat 4,734,693 judul [1], yang diantaranya terdapat judul TV series, film pendek, documenter dan sebagainya. Perkembangan industri film juga berkembang pesat. Seiring dengan berkembangnya teknologi yang digunakan dalam pembuatan film. Tidak hanya dari segi plot cerita, film saat ini harus memiliki sisi visual yang membuat konsumen terkgagum saat menontonnya. Saat ini menonton film merupakan suatu hiburan alternatif dalam mengusir kebosanan. Tidak sedikit juga seseorang menonton film karena memang hobi. Hal ini menuntut industri film untuk menghadapi

persaingan ketat dalam menciptakan terobosan baru guna memenuhi kebutuhan konsumen yang semakin beragam.

Dewasa ini, website yang berisi video dan komunitas film, seperti Netflix, MovieLens, Youtube dan sebagainya, sudah sangat populer saat ini. Pada *website* tersebut memiliki banyak *user* dan kemudian *user* tersebut bisa memberi penilaian terhadap film yang tersedia. Melihat *review* terlebih dahulu merupakan salah satu cara untuk mengetahui kualitas dari film tersebut. Namun, banyaknya jumlah *user* yang memberi *review* berbeda-beda pada suatu film membuat pembaca kebingungan dalam menyimpulkan *review* tersebut.

Selera setiap orang pasti berdeda. Seseorang bisa menyukai film berdasarkan genre, aktor atau rumah produksi. Hal ini yang menjadi permasalahan seseorang dalam menentukan film yang sesuai dengan ekspektasi. Mengingat jumlah film yang begitu banyak dan beragam jenisnya, seseorang tentu tidak memiliki cukup waktu untuk memeriksa sinopsis atau trailer satu per satu. Belum lagi jika ada film baru yang belum diketahui judulnya. Maka dari itu harapan seseorang adalah menginginkan rekomendasi film yang sesuai harapan dari berbagai aspek dengan efektifitas waktu yang maksimal.

Secara umum proses pemberian rekomendasi terdiri dari tiga langkah, yaitu: penemuan similar user, pembuatan ketetanggaan (*neighborhood*), dan penghitungan prediksi berdasarkan tetangga yang dipilih [5].

Maka aplikasi yang bisa memberikan rekomendasi film kepada pengguna sangat diperlukan untuk mendapatkan rekomendasi film yang sesuai dengan keinginan. *Content-based filtering* dan *Collaborative filtering* adalah pendekatan yang paling umum untuk membangun sebuah sistem rekomendasi [2].

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. Selain rating, CF merupakan proses penyaringan atau pengevaluasian item menggunakan opini orang lain [6].

Adapun kekurangan pada pendekatan ini yaitu tidak dapat merekomendasikan item baru kepada user apabila item tersebut belum pernah diberi rating oleh user manapun.

Untuk proses perhitungan pencarian similarity menggunakan metode *weighted sum* agar dapat memberikan rating prediksi pada item film yang belum pernah di rating oleh user aktif.

Untuk memaksimalkan hasil rekomendasi dilakukan pengelompokan film berdasarkan beberapa atribut dengan menggunakan metode *k-Nearest Neighbors* (kNN). Metode kNN melakukan klasifikasi terhadap item berdasarkan data latih dari item lain yang memiliki jarak terdekat. Dengan menggunakan dua metode tersebut hasil akhir yang didapat akan sesuai dengan harapan yaitu baik dari segi kualitatif maupun kuantitatif.

II. PENELITIAN TERKAIT

Affriantari Rochmah (2010) melakukan pengembangan fitur rekomendasi pada website solo movie untuk mendapatkan hasil rekomendasi kepada member website solo movie. Algoritma rekomendasi yang digunakan yaitu algoritma apriori untuk mendapatkan klasifikasi suatu film.

Ananda Riyandwana, dkk (2012) melakukan pengembangan Sistem Rekomendasi pada perpustakaan dalam penentuan rekomendasi buku berdasarkan karakteristik pengguna perpustakaan melalui buku yang dipinjam. Pengembangan sistem ini dilakukan dengan menggunakan metode Self Organizing Map (SOM) Clustering. Penerapan system diterapkan pada Bapersip Provinsi Jawa Timur.

Arie Anggono, dkk (2014) melakukan perekomendasi pada portal lowongan kerja dengan pendekatan metode SAW. Karena menurut mereka metode SAW adalah metode yang mampu memecahkan masalah dengan cara memberi nilai kecocokan berdasarkan CV pelamar terhadap lowongan kerja yang telah dipasang sebelumnya oleh pihak perusahaan, yang nantinya dapat membuat keputusan dalam menentukan bobot bagi setiap attribute/parameter. Selanjutnya hasil berupa skor total yang diperoleh dengan menjumlahkan seluruh hasil perkaian antara bobot dan nilai rating yang telah dinormalisasi sebelumnya

III. HASIL PENELITIAN DAN PEMBAHASAN

A. Collaborative Filtering (CF)

Adomavicius & Tuzilin (2005) menyatakan *system collaborative-based recommendation* adalah metode yang digunakan untuk memprediksi kegunaan item berdasarkan penilaian pengguna sebelumnya. Metode ini merekomendasikan item-item yang dipilih oleh pengguna lain dengan kemiripan model item dari pengguna saat ini (Sebastia, L et al., 2009).

Ketika melakukan rekomendasi, pertama kali algoritma CF akan mengidentifikasi selera user berdasarkan data rating user tersebut sebelumnya, kemudian merekomendasikan selera yang sama terhadap user aktif (Balabanovi, M, dkk, 2001).

Algoritma *collaborative filtering* terbagi atas dua kelas yang berbeda berdasarkan teori dan kepraktisannya, yaitu algoritma probabilistik dan algoritma non-probabilistik. Suatu algoritma dianggap probabilistik apabila algoritma tersebut mewakili distribusi probabilitas saat melakukan perhitungan terhadap rating atau daftar ranking item yang direkomendasi. Algoritma non-probabilistik dengan istilah lainnya yaitu *nearest neighbours algorithm*. Algoritma ini dibagi kedalam dua kelas yaitu *user-based* dan *item-based*[3].

1) *User-based nearest neighbor algorithm*: merupakan algoritma yang menggunakan teknik statistika untuk menemukan sekumpulan pengguna, dikenal sebagai tetangga (*neighbor*) yang memiliki pendapat sama atau setuju terhadap pengguna yang menjadi sasaran. Setelah sekumpulan tetangga terbentuk, sistem menggunakan algoritma yang berbeda untuk menggabungkan kesukaan *neighbours* untuk menghasilkan suatu prediksi atau rekomendasi N-teratas untuk pengguna aktif [4].

2) *Item-based collaborative filtering*: merupakan metode rekomendasi yang bekerja berdasarkan adanya kesamaan antara pemberi rating terhadap item yang dituju. Dari tingkat kesamaan item, kemudian dibagi berdasarkan parameter kebutuhan pelanggan untuk memperoleh nilai kegunaan item. Item yang memiliki nilai tertinggi maka akan dijadikan rekomendasi (Purwanto, 2009). Metode ini merupakan solusi pada beberapa masalah pada *user-based collaborative filtering* yaitu pada masalah keterbatasan (*sparsity*) dan skalabilitas serta masalah waktu dan penggunaan memori.

Pada metode ini akan diketahui nilai kesamaan antar item dengan tingkat persebaran rating kecil dan nilai kesamaan antar item cenderung lebih jarang berubah dibandingkan dengan nilai kesamaan antar pengguna. *Item-based collaborative filtering* melakukan kesamaan dengan bentuk suatu model kesamaan secara offline, maka secara otomatis akan menghemat waktu dan penggunaan memori untuk perhitungan saat pengguna mengakses halaman situs.

B. Tahapan Metode Perhitungan

Tahapan metode *collaborative filtering* terbagi menjadi 2 tahapan yaitu : perhitungan kemiripan dan perhitungan prediksi[4].

1) *Perhitungan kemiripan (Similarity)*: pada tahapan ini akan melakukan perhitungan kemiripan antar item. Untuk menghitung kemiripan dibutuhkan data rating user dimana data diperoleh secara eksplisit, user memberikan rating pada beberapa film dengan skala 1 – 5. Langkah awal adalah menentukan film yang belum pernah di rating oleh user, sehingga didapat data film yang belum di rating user untuk dicari nilai kemiripannya. Berikut adalah rumus perhitungan mencari kemiripan antar item film.

$$sim(k,l) = \frac{\sum_{u=1}^m (R_{u,k} - \bar{R}_u)(R_{u,l} - \bar{R}_u)}{\sqrt{\sum_{u=1}^m (R_{u,k} - \bar{R}_u)^2} \sqrt{\sum_{u=1}^m (R_{u,l} - \bar{R}_u)^2}} \quad (1)$$

Dimana:

Sim(k,l) adalah nilai similarity antara item k dan item l.

m adalah jumlah total user yang merating item k dan l.

\bar{R}_u adalah rating rata-rata user u.

$R_{u,k}$ dan $R_{u,l}$ adalah rating yang diberikan user u kepada item k dan l.

2) *Perhitungan prediksi (Similarity)*: setelah didapatkan nilai kemiripan, langkah selanjutnya adalah menghitung nilai prediksi dengan mengurutkan hasil nilai kemiripan dari nilai terbesar ke kecil. Kemudian hitung nilai prediksi menggunakan persamaan weight sum.

$$P_{(u,k)} = \frac{\sum_{i=1}^n (R_{u,i} * S_{i,j})}{\sum_{i=1}^n |S_{i,j}|} \quad (2)$$

Dimana:

P(u,k) adalah nilai prediksi antara item u dan item k.

$R_{u,i}$ adalah nilai rating user pada item i.

$S_{i,j}$ adalah nilai similarity dari hasil perhitungan persamaan ke 1 untuk item i dan j.

C. Pengujian Perhitungan

1) *Pencarian User serupa*: Untuk melakukan pencarian user serupa, langkah awal yang harus dilakukan adalah memindahkan database rating kedalam matriks 2 dimensi. Pada database rating terdapat satu buah *rating user* terhadap beberapa film.

Rating pada film yang telah di isi oleh user memiliki nilai antara 1-5. Sedangkan untuk user yang belum pernah memberi rating kepada suatu film maka rating akan diberi nilai 0. Uji coba perhitungan menggunakan tiga data user yang telah mengisi beberapa rating dari sepuluh film. Maka didapat tabel matriks (Tabel 1) sebagai berikut.

TABEL I
TABEL MATRIKS RATING FILM

DATA RATIN G	Judul Film										
	A	B	C	D	E	F	G	H	I	J	Ru
Miftah	2	4	5	0	0	0	0	0	4	5	4
Arif	2	5	0	4	3	3	2	2	0	0	3
Hanafi	2	1	0	2	3	0	3	0	1	1	2

Berdasarkan tabel di atas, Akan dicari nilai prediksi pada film D dengan Miftah sebagai *user* aktif. Didapat user

Arif dan Hanafi sebagai user serupa untuk melakukan perbandingan film A dan film B terhadap film D.

2) *Perhitungan Nilai Similarity Antar Film*: Selanjutnya akan dicari film rekomendasi untuk user Miftah dengan melakukan perhitungan menggunakan persamaan(1) maka didapat Tabel 2 yang merupakan nilai kemiripan sebagai berikut.

TABEL II
HASIL NILAI KEMIRIPAN MENGGUNAKAN PERSAMAAN (1)

Tabel Kemiripan	
Sim(A,D)	1
Sim(B,D)	0.4

Berdasarkan tabel di atas nilai kemiripan antara film A dan D didapat 1, kemudian kemiripan antara film B dan D didapat 0.4

3) *Menghitung Nilai Prediksi Rating Film*: Setelah didapat nilai kemiripan antar film, selanjutnya menghitung nilai prediksi menggunakan persamaan(2) maka didapat nilai prediksi sebagaimana ditampilkan pada Tabel 3

TABEL III
NILAI PREDIKSI MENGGUNAKAN PERSAMAAN (2)

Tabel Prediksi	
P(Miftah,D)	2.571

Dari hasil perhitungan diatas bahwa hasil prediksi rating user Miftah terhadap film D adalah 2.571. Maka, bisa disimpulkan hasil perhitungan prediksi *rating* Miftah terhadap film D adalah 2.571 dengan skala rating 1 sampai 5.

4) *Sorting*: Jika semua film dilakukan perhitungan prediksi rating, maka akan diperoleh data sebagai berikut yang ditampilkan pada Tabel 4

TABEL IV
HASIL PERHITUNGAN PREDIKSI RATING

Tabel Prediksi	
P(Miftah,D)	2.571
P(Miftah,E)	1.432
P(Miftah,F)	4.233
P(Miftah,G)	2.332
P(Miftah,H)	0.353

Berdasarkan data tabel di atas, untuk menampilkan hasil rekomendasi terbaik maka dilakukan pengurutan nilai prediksi rating secara *descending*

IV. KESIMPULAN

Aplikasi dapat memberikan rekomendasi terhadap user aktif dengan cara menghitung rating prediksi dengan kondisi beberapa user lainnya sudah pernah memberi rating pada film yang akan di prediksi untuk user aktif.

User aktif yang memiliki banyak film yang akan dihitung prediksi ratingnya, maka dengan perhitungan menggunakan metode *weight sum* dapat membuat urutan

hasil prediksi dari rating terbesar ke rating terkecil untuk urutan perekomendasi film.

Untuk Penelitian selanjutnya dibutuhkan metode tambahan untuk menutupi kekurangan pada prediksi rating dimana kondisi film belum pernah di rating sama sekali oleh seluruh user.

REFERENSI

- [1] "Press Room," IMDB, 2018. [Online]. Available: <https://www.imdb.com/pressroom/about/>.
- [2] Adomavicius, Gediminas dan Tuzhilin, Alexander (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, June 2005.
- [3] Schafer, B., Joseph, Konstan and John Riedl, "Recommender Systems in E-Commerce," in Minneapolis, University of Minnesota, 2007.
- [4] B. Sarwar, "Item based collaborative filtering recommendation algorithms" ACM.
- [5] Akhiro, R. "Studi Multi Criteria Decision Making (MDCDM) untuk Recommender System Bursa Tenaga kerja". Institut Teknologi Bandung, 2008.
- [6] Schafer, J.B., Frankowski, D., Herlocker, J. dan Sen, S. "Collaborative Filtering Recommender System". Springer-Verlag, Berlin, Heidelberg, 2007.