

# Implementasi Machine Learning Untuk Rekomendasi Film Di Imdb Menggunakan Collaborative Filtering Berdasarkan Analisa Sentimen IMDB

<sup>1</sup>Maylida Izattul Wardah, <sup>2</sup>Syahrizal Dwi Putra

Program Studi Teknik Informatika, Fakultas Ilmu Komputer,  
Universitas Esa Unggul  
Jl. Arjuna Utara No.9 Duri Kupa, Kec. Kb. Jeruk, Kota Jakarta barat

**Email:** [maylidaww@gmail.com](mailto:maylidaww@gmail.com) , [syahrizal.dwi@esaunggul.ac.id](mailto:syahrizal.dwi@esaunggul.ac.id)

**Received:** May 26, 2022, **Revised:** June 28, 2022, **Accepted:** July 25, 2022

## Abstrak

Perkembangan industri perfilman di dunia sangat berkembang pesat. IMDB adalah website yang telah lama digunakan untuk menyajikan informasi dan berbagi opini antar penikmat film yang ada di seluruh dunia, tanggapan mereka menjadi tolak ukur kesuksesan dari sebuah film terdapat banyak judul film, yang menyebabkan penonton kesulitan dalam memilih film. Meskipun dengan klasifikasi genre yang telah ada di website IMDB, tetapi klasifikasi ini terbilang cukup umum karena masih belum dapat mengatasi preferensi genre dari penonton yang berbeda-beda, seperti penonton dengan preferensi film yang memiliki genre comedy dengan sedikit bumbu romance, tentu akan berbeda hasilnya dengan film dengan genre comedy dengan bumbu action, banyak jenis genre film yang siap ditonton, tetapi banyak pengguna atau user yang kebingungan film apa yang akan ditonton. Oleh karena itu dibutuhkan sistem untuk merekomendasikan film agar sesuai dengan yang pelanggan atau user inginkan. Pada penelitian ini, penulis menerapkan teknologi komputersasi dalam menganalisa rekomendasi film dengan Collaborative Filtering yaitu metode User-based Filtering dan Item-based Filtering untuk mencari kesamaan dan memprediksi item dan juga analisa sentimen untuk melihat hasil ulasan dan mengklasifikasi komentar dari komentar pengunjung IMDB menggunakan Naive Bayes Classifier

**Kata kunci:** perfilman, metode User-based Filtering dan Item-based Filtering, Naive Bayes Classifier.

*Abstract – The development of the film industry in the world is growing rapidly. IMDB is a website that has long been used to provide information and share opinions among moviegoers around the world, they are a benchmark for the success of a film. Even with the genre classification that already exists on the IMDB site, this classification is quite general because it still cannot cope with the genre preferences of different audiences, such as viewers with film preferences that have a comedy genre with a little romance spice, of course the results will be different from those of the audience. films with the comedy genre with action spices, many types of films are ready to be watched, but many users or users confuse what films to watch. Therefore, a system is needed to recommend films to suit the customer or what the user wants. In this study, the authors apply computerized technology in analyzing film recommendations with Collaborative Filtering, namely User-based Filtering and Item-based Filtering methods to find similarities and predict items and also analyze senders to see the results of reviews and classify comments from IMDB visitors using Naive Bayes Classifier*



This work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/).  
<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>

**Keywords:** *the film, User-based Filtering and Item-based Filtering methods, Naive BayesClassifiers*

## 1 Pendahuluan

Film merupakan salah satu dunia hiburan yang sangat digemari oleh masyarakat di Indonesia. Dikarenakan dengan semakin ramainya penggemar film, baik film Asia maupun film Barat. Maka timbul pemikiran untuk mengembangkan informasi seputar dunia industri perfilman IMDb ([www.imdb.com](http://www.imdb.com)) Situs film yang memiliki database film yang lengkap dan menyediakan beberapa fasilitas, antara lain untuk pencarian film dan dapat juga memberikan rekomendasi film. Rekomendasi film yang diberikan oleh IMDb adalah rekomendasi film berdasarkan keterkaitan film yang disukai oleh penonton

Permasalahan yang muncul terkadang para penggemar film mengalami kebingungan saat mencari dan mendapatkan informasi film di internet. Penggemar film terkadang mencari informasi film di beberapa situs untuk dibandingkan, terutama pada bagian rekomendasi film, padahal rekomendasi yang dihasilkan terkadang berbeda.

Berdasarkan permasalahan yang ada penelitian ini dilakukan untuk menciptakan sebuah rekomendasi dengan Collaborative Filtering berdasarkan Sentiment Analysis IMDb. rekomendasi akan memunculkan film yang dipilih oleh user atau rekomendasi dari genre film yang sering dilihat berdasarkan nilai rating dan ulasan yang sudah ada.

### 1.1 Identifikasi Masalah

Berdasarkan Latar Belakang maka permasalahan yang akan diidentifikasi untuk penelitian ini yaitu:

1. Bagaimana penerapan metode Collaborative Filtering dalam sebuah sistem rekomendasi ?
2. Bagaimana cara menilai kecocokan rekomendasi film menggunakan User-based filtering dan Item-based filtering ?
3. Bagaimana tingkat akurasi film menggunakan metode MAE dan RMSE ?
4. Bagaimana melakukan analisa sentiment berdasarkan komentar pengunjung film ?

### 1.3 Tujuan Tugas Akhir

Adapun tujuan penulis membuat penelitian ini selain untuk memenuhi Tugas Akhir yaitu:

1. Membangun sistem rekomendasi dengan metode Collaborative Filtering
2. Mengetahui tingkatan kecocokan rekomendasi film yang dibuat dengan User-based Filtering dan Item-based Filtering
3. Mengetahui akurasi sistem rekomendasi film menggunakan Algoritma MAE dan RMSE
4. Melakukan analisa sentiment berdasarkan komentar pengunjung film

### 1.4 Adapun tujuan penulis membuat penelitian ini selain untuk memenuhi Tugas Akhir yaitu:

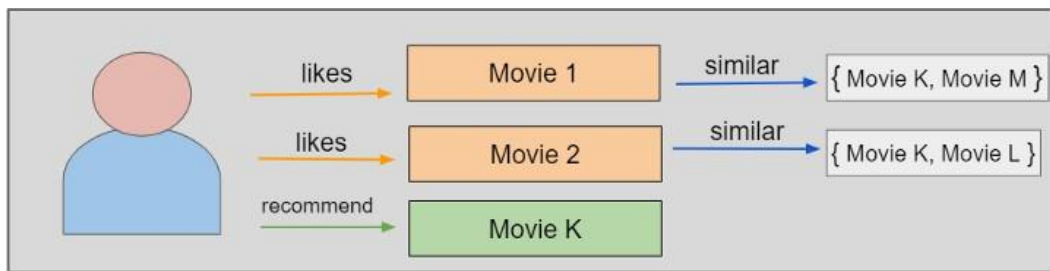
1. Mengetahui tingkatan kecocokan sistem rekomendasi film yang dibuat dengan review film yang dibaca user
2. Membangun sistem rekomendasi dengan
3. Mengetahui akurasi sistem rekomendasi film menggunakan metode Pearson Correlation
4. Mengetahui klasifikasi komentar positif dan negatif

## 2 Tinjauan Literatur (or Literature Review)

### 2.1 Sistem Rekomendasi



This work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/).  
<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>



Gambar 1: Contoh sistem rekomendasi film pemfilteran kolaboratif

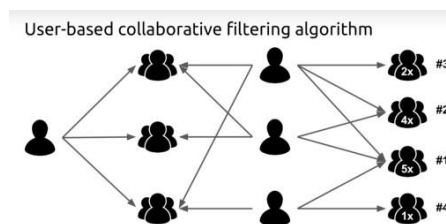
Sistem Rekomendasi adalah suatu program yang melakukan prediksi sesuatu item, seperti rekomendasi film, musik, buku, berita dan lain sebagainya yang menarik user (Scafer et al., 2001).

rekomendasi film, musik, buku, berita dan lain sebagainya yang menarik user (Scafer et al., 2001). Sistem ini berjalan dengan mengumpulkan data dari pengguna secara langsung maupun tidak. Sistem ini awalnya merupakan metode penyaringan informasi atau information filtering, sebuah metode untuk menyaring informasi sebagai akibat dari terlalu banyaknya informasi di internet dan media (Hidayat, 2007).

## 2.2 Collaborative Filtering

Collaborative Filtering merupakan sebuah metode dalam membuat prediksi dengan cara menyaring informasi item dari opini orang lain. Fokus utama dalam sistem rekomendasi collaborative filtering adalah untuk memanfaatkan riwayat opini pengguna aktif lain untuk memprediksi item yang mungkin akan disukai atau diminati oleh seorang pengguna. Implementasi yang paling sederhana dari pendekatan ini adalah membuat rekomendasi kepada pengguna aktif berdasarkan item yang disukai pengguna ;lain dengan riwayat selera yang serupa(Ricci et.al.,2011)

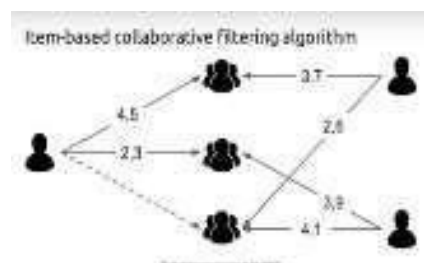
### 2.1.1 User-based Filtering



Gambar 1. Algoritma User-Based Collaborative Filtering

User-based filtering mengasumsikan bahwa cara untuk menemukan item yang menarik bagi user tertentu adalah dengan mencari user lain yang memiliki minat yang sama. Jadi, pada awalnya User-based filtering mencoba untuk menemukan user lain berdasarkan pada user similarity dan kemudian setiap nilai rating dari user lain akan dijadikan bahan rekomendasi bagi user aktif (Yu, 2015).

### 2.2.2 Item-based Filtering



Gambar 2. Algoritma Item-Based Collaborative Filtering



Tujuan dari Item-based filtering adalah mencari pola pemberian rating terhadap sebuah item dan kemudian mencoba memprediksi rating yang akan diberikan seorang pengguna terhadap item lain. Contohnya Eric menyukai film “Forrest Gump” dan “Wall-E”. Namun dia belum menonton film. “Titanic”. Maka Eric melihat bahwa Justin dan Kevin sudah memberikan peringkat yang sama pada film “Titanic” dengan dua film sebelumnya, maka Eric menyimpulkan bahwa dia juga akan menyukai film “Titanic” (Ricci et al., 2011),

## 2.3 Similarity

Similarity adalah suatu metode pada machine learning yang berfungsi untuk menghitung kemiripan antara 2 data atau lebih. Perhitungan similarity menggunakan metode algoritma seperti Pearson correlation, cosine similarity, Jaccard similarity, dan masih banyak lagi.

### 2.3.1 Pearson Correlation

Pearson Correlation adalah salah satu algoritma yang biasa digunakan dalam menghitung similarity antara user satu dengan user lainnya. Korelasi (correlation) adalah sebuah teknik pengukuran yang menentukan kedekatan hubungan antara dua himpunan bilangan yang berbeda. Perhitungan korelasi memiliki syarat dimana himpunan bilangan yang dihitung harus memiliki urutan yang tetap dan berpasangan satu dengan lainnya antar kedua himpunan

## 3 Metode Penelitian (or Research Method)

Metode Penelitian adalah langkah atau proses dalam penelitian seperti pengumpulan data yang dibutuhkan, wawancara dan observasi yang digunakan, beberapa metode yang dapat membantu dalam penulisan ini. Diantaranya adalah sebagai berikut:

- a. Menentukan Topik  
Pada tahap ini penulis akan menentukan topik yang akan dikerjakan oleh penulis, dan juga menentukan apa saja kebutuhan yang diperlukan untuk membantu penulis untuk menyelesaikan pengerjaan tugas akhir ini.
- b. OBSERVASI  
Teknik observasi dilakukan dengan cara datang secara online ke website IMDB yang akan dikunjungi, untuk mengamati objek-objek penelitian yang ingin diteliti.
- c. Web Scrapping  
Setelah melakukan observasi, penulis juga melakukan scrapping Data dari website IMDb.com menarik beberapa data yang akan dibutuhkan. Hasil dari proses scrapping ini ke dalam format CSV atau xlsx dengan begitu data dapat mudah digunakan untuk keperluan analisis data
- d. Studi Kepustakaan  
Tahapan ini melakukan penggalan data dan pengumpulan informasi dengan cara mempelajari serta memahami sebuah literatur yang berkaitan dengan penelitian ini berupa buku dan jurnal artikel, agar menjadi bahan referensi untuk pembuatan tugas akhir ini.
- e. Metode Analisis Data  
Peneliti melakukan analisis data menggunakan metode user-based dan item-based filtering untuk rekomendasi film serta sentiment analysis untuk mengetahui nilai ulasan dari pengguna. Collaborative Filtering digunakan untuk mengetahui prediksi rating film pada semua pengguna dan juga Analisis Sentimen digunakan untuk mengklasifikasi komentar user menggunakan naive bayes classifier
- f. Hasil dan Pembahasan (or Results and Analysis)  
Proses collaborative filtering ditanamkan pada salah satu laman dari sistem rekomendasi ini dan disimpan dalam basis data guna mendapatkan hasil yang otomatis, kemudian hasil tersebut akan dipanggil dalam bentuk tabel matrix. Untuk dapat dilakukan perhitungan, dibutuhkan beberapa data sampel 5 objek wisata dan pengguna yang memberikan rating di objek wisata, seperti terlihat pada tabel 1:



Tabel 1. Rating pengguna

No	Nama Film	Pengguna 1	Pengguna 2	Pengguna 3	Pengguna 4	Pengguna 5
1	Minions: The Rise of Gru	5	4		5	4
2	The Walking Dead			4	4	4
3	Blue Bloods		3		3	4
4	La casa de papel	4	2		4	
5	Friends			5	4	4

Algoritma collaborative filtering selanjutnya akan melakukan perhitungan tingkat kemiripan setiap objek wisata dengan menggunakan metode pearson correlation. Input data yang digunakan adalah rating dari pengguna. Output dari perhitungan tersebut berupa nilai kemiripan dari objek film pertama dan film kedua dibandingkan. Sehingga dihasilkan:

Tabel 2. Hasil perhitungan kemiripan

Pengguna	Film 1	Film 2	Kemiripan
User 1 User 2	Minions: The Rise of Gru	Blue Bloods	-1,25
User 1 User 4	Minions: The Rise of Gru	Blue Bloods	0,75
...	...	...	...

Setelah dilakukan proses kemiripan, metode weight sum dilakukan untuk mencari kemiripan dari objek film dan rating dari objek film tersebut. Output dari perhitungan tersebut akan menghasilkan prediksi nilai rating dari seorang pengguna terhadap objek tertentu film dengan menggunakan persamaan (2.4), dimana nilai prediksi tertinggi akan berada di posisi pertama, seperti yang didapat pada tabel.

Tabel 3 Prediksi rating

Pengguna	Objek	Prediksi
User 4	Minions: The Rise of Gru	4
User 4	Blue Bloods	4
User 4	La casa de papel	3
User 2	La casa de papel	3

Tahapan selanjutnya adalah melakukan proses sentimen analisis untuk mengetahui nilai dari ulasan dari pengguna. Tahapan awal dari dalam sentiment analisis akan yang digunakan pada K-Nearest neighbors adalah text preprocessing, terdapat tiga tahapan preprocessing: tokenizing, filtering, dan stemming. Proses tokenizing merupakan proses pemotongan string input berdasarkan tiap kata yang menyusunnya. Selain itu, spasi digunakan untuk memisahkan antar kata tersebut, seperti ditunjukkan pada tabel 4.



Tabel 4. Proses sentiment analisis sebelum dan sesudah

N o	Dokum en	Hasil text preprocessing
1	Nothing too impressive or new.. Te scenes look contrived. Learning kung-fu scenes were an utter waste of time. Too slapstick even for kids. Hope there will be a better storylinenext time.	Nothing too impressive.. Te scenes look contrived. Learning kung-fu scenes were an utter waste of time. Too slapstick even for kids
2	The stories, the characters, the way it changes from fighting zombies to fighting humans and back to zombies. Darryl, Michone and of course Rick are my favourite	The stories, the characters, the way it changes fromfighting zombies to fighting humans and back to zombies.
3	I don't like the show anymore. Gone full woke. It's a reflection of society and society sucks. Cancel the show.	I don't like the show Gone fullwoke reflection of society and society sucks.
4	Female inspectors is dumb . Robber keep coming front ofher she don't use her brain. . Don't waste your time with this series I'm in season 1 episode 3 I'm already bored.	Female inspectors is dumb . Robber keep coming front ofher she don't use her brain. . Don't waste your time withthis series
5	Even in 2019 it is very funny. A really great show. Mustwatch	2019 it is very funny. A really great show. Must watch

Dari tabel tersebut, terdapat proses tokenizing dengan menghilangkan karakter selain ‘a’ sampai dengan ‘z’ yang dianggap sebagai delimiter dan mengubah kalimat tersebut menjadi potongan string kata penyusunnya. Selanjutnya proses filtering dijalankan untuk mengeliminasi kata-kata yang kurang penting seperti kata ‘untuk’ pada ulasan nomor 1, kata ‘disini’, ‘selain’, ‘itu’, ‘agak’, pada ulasan lainnya. Dan dilanjutkan dengan proses stemming guna mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (dasar) dengan menggunakan aturan-aturan tertentu, seperti menghilangkan imbuhan pada kata ‘tempat’ menjadi ‘tempat’, ‘menyebabkan’ menjadi ‘sebab’ dan lain sebagainya..

#### g. Kesimpulan

Proses perhitungan rekomendasi objek pariwisata menggunakan metode collaborative filtering berhasil dilaksanakan dan diimplementasikan pada objek film. Berdasarkan hasil perhitungan rating, setiap pengguna dapat melihat rekomendasi objek film yang terbaik dan layak untuk dikunjungi terlebih dahulu. Begitu pula dengan melihat hasil ulasan menggunakan sentimen analisis, sistem rekomendasi memberikan hasil yang terbaik dari perhitungan sentiment analisis menggunakan k-nearest neighbor. Kedua metode ini akan mendapatkan hasil yang lebih baik, jika objek wisata yang dijadikan sebagai variabel diperbanyak sebanyak mungkin dan setiap objek film mendapatkan rating-rating dari setiap pengguna. Begitu juga dengan prinsip sentiment analisis, makin banyak ulasan yang diolah maka sistem akan merekomendasikan objek film terbaik.

#### Referensi (Reference)



This work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/).

<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>



DOI: <https://doi.org/10.52362/jmijayakarta.v2i3.868>

- [1] Albertus Bayu Aji Priyono (2016). Performa Apriori dan Collaborative Filtering untuk Sistem Rekomendasi (Studi Kasus: Rumah Sakit Yukum Medical Centre). *Jurnal Ilmiah Informatika dan Komputer*
- [2]. Muhammad Fajriansyah, dan Putra Pandu Adikara, dan Agus Wahyu Widodo (2016). Sistem Rekomendasi Film Menggunakan Content Based Filtering (Studi Kasus: IMDb). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6, 2188-2199
- [3]. Ampazis N. “.Collaborative Filtering via Concept Decomposition on the Netflix Dataset “, ECAI Workshop on Recommender Sistem. Pp. 26-30, Greece 2008
- [4]. Tren Positif Film Indonesia | Indonesia.go.id (2019). Available at: <https://indonesia.go.id/ragam/seni/sosial/tren-positif-film-indonesia> (Accessed: 28 Desember 2021)
- [5]. Witta Listiya Ningrum. Pendekatan Deep Learning untuk Sistem Rekomendasi Review Film. <https://docplayer.info/195391461-Pendekatan-deep-learning-untuk-sistem-rekomendasi-review-film.html>
- [6]. Reddy, S. R. S. et al. (2019). ‘Content - Based Movie Recommendation System Using Genre Correlation’, in Satapathy, S. C., Bhateja, V., and Das, S. (eds) *Smart Intelligent Computing and Applications*. Singapore: Springer Singapore, pp. 391 – 397
- [7]. Badan Pusat Statistik. (2012, Februari) BPS. [Online]. Available: <http://www.bps.go.id/index.php/istilah/357>. Collaborative Filtering: Creating the Best Teams Ever. (2017, Februari) Maurits van der Goes [Online]. Available: <https://neo4j.com/blog/collaborative-filtering-creating-teams/>



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).  
<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta>