

1. PENDAHULUAN

1.1. Latar Belakang

Pada saat ini, dunia hiburan sedang meningkat pesat di dunia. Salah satunya yaitu film. Dikarenakan banyaknya film yang selalu bertambah setiap harinya, pengguna harus memilih film yang ingin ditonton dengan opsi yang sangat banyak [1]. Hal ini merupakan masalah yang dihadapi oleh kebanyakan pengguna. Dengan banyaknya informasi dan pilihan yang ada, pengguna akan sulit untuk menemukan film yang sesuai dengan selernya. Pengguna yang berbeda pastinya memiliki selera yang berbeda juga. Maka dari itu, sangat penting untuk pengguna memiliki suatu sistem rekomendasi untuk memudahkan pengguna dalam mempersempit film berdasarkan ketertarikan dari masing-masing pengguna [2].

Sistem rekomendasi film merupakan proses untuk memfilter apa yang diinginkan pengguna dengan mempersempit opsi-opsi yang dimiliki dan memberikan rekomendasi film sesuai dengan minat dan selera masing-masing pengguna [2]. Telah banyak penelitian yang telah membuat rekomendasi film dengan pendekatan yang berbeda-beda. Dan setiap pengguna memiliki ketertarikan yang berbeda-beda tergantung pada beberapa faktor, misalnya ada pengguna yang memiliki ketertarikan berdasarkan genre film, berdasarkan rating, ataupun berdasarkan dari sinopsis film tersebut [3]. *Collaborative filtering* dan *content-based filtering* merupakan dua metodologi yang sering digunakan dalam sistem rekomendasi. *Content-based filtering* merupakan penyaringan konten menggunakan informasi-informasi yang didapatkan dan memberikan rekomendasi berdasarkan profil atau interaksi yang dilakukan oleh pengguna [4], sedangkan *collaborative filtering* merupakan penyaringan berdasarkan pengguna yang memiliki minat yang sama dan memberikan rekomendasi berdasarkan semua pengguna yang serupa [5].

Pada penelitian ini, pendekatan yang digunakan yaitu *content-based filtering* dengan memanfaatkan algoritma TF-IDF (*Term Frequency - Inverse Document Frequency*) dan *cosine similarity*. Pendekatan ini digunakan karena dapat menebak apa yang disukai pengguna dan memberikan rekomendasi

berdasarkan konten yang didapatkan dari pengguna tersebut berupa riwayat atau interaksinya di masa lampau. Data yang akan digunakan yaitu data yang tersedia untuk umum (*MovieLens*). Dalam *dataset* tersebut terdapat beberapa informasi tentang film. Salah satu informasi tersebut berupa sinopsis yang biasanya terdapat dalam film. Penulis membandingkan model rekomendasi berdasarkan genre, *keyword*, dan kombinasi dari kedua variabel tersebut. Sinopsis ini akan di ekstrak menjadi *keyword* dengan menggunakan *library* Python, Gensim. Kemudian, *keyword* akan diproses dengan melakukan ekstraksi fitur yang memanfaatkan TF-IDF. TF-IDF digunakan untuk mengubah setiap kata yang ada pada dokumen menjadi nilai numerik [6]. Dari nilai numerik tersebut, akan dihitung kemiripannya dengan menggunakan *cosine similarity*. Hasil dari perhitungan inilah yang akan memberikan rekomendasi dengan melihat *top-k similarity* dengan nilai kemiripan tertinggi.

1.2. Perumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah yang bisa diambil yaitu :

1. Bagaimana cara membuat sistem rekomendasi menggunakan *content-based filtering*.
2. Bagaimana mengimplementasikan *keyword extraction* pada teks sinopsis film.
3. Bagaimana mengimplementasikan dua algoritma yaitu TF-IDF untuk *feature extraction* dan *cosine similarity* untuk menghitung kemiripannya.
4. Bagaimana melakukan analisa terhadap performansi model.

1.3. Tujuan

Berdasarkan rumusan masalah tersebut, maka tujuan dari penelitian ini yaitu :

1. Membuat suatu sistem rekomendasi menggunakan *content-based filtering*.
2. Mengimplementasikan *keyword extraction* pada teks sinopsis film.
3. Mengimplementasikan dua algoritma yaitu TF-IDF untuk *feature extraction* dan *cosine similarity* untuk menghitung kemiripannya.
4. Menganalisis performansi model agar mendapatkan hasil yang maksimal.

1.4. Batasan Masalah

Dalam penelitian ini, adanya Batasan masalah yang bertujuan untuk memudahkan dalam pembahasannya. Batasan-batasan yang ada dalam penelitian ini sebagai berikut :

1. Data yang digunakan diambil dari *dataset* yang tersedia untuk umum (*MovieLens*) berbahasa inggris.
2. Algoritma yang digunakan dalam membuat sistem rekomendasi ini yaitu TF-IDF untuk *feature extraction* dan *cosine similarity* untuk menghitung kemiripannya.
3. Model yang dihasilkan yaitu *Top-K Similar Item*.

1.5. Metodologi Penelitian

Rencana kegiatan yang akan dilakukan pada penelitian ini adalah sebagai berikut :

1. Studi Literatur
Pada tahap ini akan dilakukan review terhadap paper yang terkait dengan penelitian yang akan dibuat seperti penerapan metode *content-based filtering*, algoritma TF-IDF dan *cosine similarity* untuk dijadikan landasan teori dalam tugas akhir.
2. Analisis dan Perancangan Sistem
Pada tahap ini, akan dilakukan pencarian tentang *dataset* dengan parameter apa saja yang digunakan, menganalisis sistem untuk mencari tau kelebihan dan kekurangan yang ada dalam penelitian ini, serta membuat tahapan sistem yang dibangun seperti pengumpulan data dan pembuatan flowchart.
3. Implementasi Sistem
Implementasi dilakukan berdasarkan sistem yang telah dibuat, yaitu pembuatan sistem rekomendasi dengan menggunakan algoritma TF-IDF, dan *cosine similarity*.
4. Analisis Hasil Implementasi

Pada tahap ini, akan dilakukan analisis hasil dari sistem yang telah dibuat. Apakah hasil sudah maksimal atau harus ada pengembangan sistem kembali dengan mengevaluasi beberapa hal yang harus diperbaiki.

5. Penulisan Laporan

Seluruh kegiatan yang telah dilakukan akan dituliskan di dalam sebuah laporan.