1. Pendahuluan

Latar Belakang

Konten digital yang ada di Internet saat ini sangat berkembang. Ini memperumit proses memperoleh informasi yang membantu dalam menyediakan *user* dengan dukungan individual dalam memisahkan sejumlah besar data yang dapat diakses [1]. Dalam era digital saat ini, industri hiburan telah mengalami perkembangan pesat berkat kemajuan teknologi informasi dan internet. Salah satu bidang yang telah mendapatkan perhatian khusus adalah industri film. Dengan munculnya platform streaming online, penonton memiliki akses lebih mudah dan fleksibel untuk menikmati berbagai film dari berbagai genre dan tahun produksi. Namun, dengan jumlah film yang semakin banyak dan beragam, *user* seringkali merasa kesulitan untuk memilih film yang sesuai dengan preferensi dan minat mereka.

Sistem rekomendasi film berfungsi sebagai solusi yang potensial untuk mengatasi masalah ini. Namun, masalah nyata dalam implementasi sistem rekomendasi sering kali muncul. Salah satu masalah adalah fenomena yang dikenal sebagai "*filter bubble*," di mana sistem hanya merekomendasikan film-film yang serupa dengan apa yang sudah pernah ditonton oleh *user*. Akibatnya, *user* cenderung terjebak dalam zona nyaman mereka dan tidak banyak mengeksplorasi film-film baru yang mungkin juga menarik bagi mereka. Fenomena ini sering kali muncul dalam layanan seperti platform media sosial seperti *Facebook*, platform streaming musik dan video seperti *YouTube*, *Netflix*, serta sistem rekomendasi produk atau konten lainnya [12].

Terdapat penelitian yang sudah berhasil menerapkan sistem rekomendasi film, seperti yang dilakukan oleh Ms. Sushmita Roy *Semi Supervised Learning* menggunakan nilai akurasi RMSE dan MAE. Mendapatkan nilai akurasi RMSE 0,4916 dan MAE 0,3453 [10]. Berikut penelitian yang sudah berhasil menerapkan sistem rekomendasi film, seperti yang dilakukan oleh Bagher Rahimpour Cami, Hamid Hassanpour, Hoda Mashayekhi menggunakan metode content-based filtering dental algoritma SVD++ menggunakan nilai akurasi *Precision dan Recall*. Dengan top N = 5 Mendapatkan nilai akurasi Recall 0,16 dan *Precision* 0,76 [11].

Dalam konteks inilah penelitian tentang pengembangan sistem rekomendasi film dengan kasus nyata menjadi sangat relevan. Dengan memerhatikan masalah-masalah nyata ini, penelitian ini bertujuan untuk mengembangkan solusi yang lebih baik, lebih personal, dan lebih akurat dalam merekomendasikan film-film yang sesuai dengan minat *user*. Dengan cara ini, penonton dapat lebih mudah menemukan film-film yang menghibur dan relevan, sambil tetap terbuka terhadap penemuan-penemuan baru yang menarik. Penelitian ini bertujuan untuk mengimplementasikan sistem rekomendasi film menggunakan metode *collaborative filtering*, salah satunya diimplementasikan dengan KNN yang diimplementasikan dengan berbagai metrik kesamaan. Kemudian implementasi ini akan dievaluasi dan dibandingkan dari segi akurasi dan efisiensi, untuk menentukan sistem mana yang lebih cocok untuk menyelesaikan permasalahan yang dihadapi.

Topik dan Batasannya

Tugas akhir ini fokus pada pemberian prediksi *rating* film sebagai rekomendasi, menggunakan metode *user-based collaborative filtering*. Evaluasi dilakukan dengan mengukur tingkat akurasi menggunakan metode MAE pada variasi model KNN. Namun, ada beberapa batasan masalah dalam penelitian ini, yaitu:

1. Data yang dianalisis hanya terbatas pada data yang ada dalam dataset yang digunakan.
2. Dataset yang digunakan berasal dari sumber yang diambil dari internet.
3. Penelitian ini akan membandingkan metode *similarity* yang berbeda.
4. penelitian ini akan mencoba menentukan jumlah tetangga dalam KNN dengan mempertimbangkan rentang nilai 10.
5. data yang digunakan dalam penelitian ini adalah data mengenai film yang diambil dari dataset *MovieLens*.

Tujuan

Tujuan dari penelitian ini adalah untuk menguji dan membandingkan performa berbagai metode *similarity* dalam algoritma KNN, termasuk *pearson correlation*, *pearson baseline*, dan *cosine*. Penelitian ini juga bertujuan untuk mengevaluasi pengaruh variasi jumlah tetangga pada hasil prediksi. Dengan melakukan perbandingan dan evaluasi ini, penelitian ini ingin menemukan kombinasi metode *similarity* dan jumlah tetangga yang paling akurat dalam memberikan prediksi *rating* untuk merekomendasikan film kepada *user*. Tujuan utamanya adalah untuk menghasilkan model prediksi yang dapat memberikan rekomendasi film kepada *user* dengan tingkat akurasi yang tinggi, sehingga meningkatkan pengalaman *user* dalam memilih film yang sesuai dengan preferensi mereka. Evaluasi akurasi dilakukan berdasarkan nilai *Mean Absolute Error (MAE)* yang digunakan sebagai ukuran performa model prediksi.