**LAPORAN AKHIR**

**Sistem Rekomendasi Film Berbasis Review**

****

**Disusun Oleh :**

**Salman Alfarisi – A11.2022.14730**

**UNIVERSITAS DIAN NUSWANTORO SEMARANG**

**2024**

# **BAB I**

# **Pendahuluan**

Dalam era digital saat ini, ulasan atau review pengguna memainkan peran penting dalam memberikan panduan bagi individu untuk membuat keputusan, terutama dalam memilih film. Review tidak hanya berfungsi sebagai sumber informasi tentang kualitas film, tetapi juga mencerminkan opini dan perasaan penonton. Dengan meningkatnya jumlah review yang tersedia di berbagai platform, seperti IMDb, Rotten Tomatoes, dan media sosial, diperlukan teknologi cerdas untuk menganalisis data teks ini secara otomatis dan memberikan rekomendasi yang relevan.

Natural Language Processing (NLP) adalah cabang kecerdasan buatan yang fokus pada pemrosesan dan analisis data teks. Dalam proyek ini, NLP digunakan untuk menganalisis ulasan pengguna, mengekstraksi fitur penting, dan mengklasifikasikan sentimen atau kategori berdasarkan isi ulasan tersebut. Metode klasifikasi K-Nearest Neighbors (KNN) digunakan untuk menentukan pola dari data teks yang telah diolah, yang dapat membantu dalam memberikan rekomendasi film yang sesuai dengan preferensi pengguna.

Pada sisi modeling, pendekatan berbasis Transformer seperti BERT (Bidirectional Encoder Representations from Transformers) atau GPT (Generative Pre-trained Transformer) menjadi pilihan utama karena performanya yang luar biasa dalam memahami konteks dan makna dalam teks. Dengan menggunakan model Transformer, proyek ini bertujuan untuk menghasilkan representasi teks yang lebih akurat, yang kemudian dapat digunakan dalam proses klasifikasi dan rekomendasi.

Proyek ini bertujuan untuk mengembangkan sistem rekomendasi film berbasis ulasan pengguna dengan memanfaatkan metode klasifikasi KNN dan model Transformer terbaik. Sistem ini diharapkan dapat memberikan rekomendasi yang lebih personal, relevan, dan efektif dengan memanfaatkan kekuatan NLP untuk memahami konteks review secara mendalam.

Manfaat dari proyek ini tidak hanya mencakup peningkatan pengalaman pengguna dalam menemukan film yang sesuai, tetapi juga memberikan kontribusi pada pengembangan teknologi analisis teks dan sistem rekomendasi yang lebih cerdas. Dengan memahami preferensi pengguna secara lebih baik, aplikasi ini dapat diimplementasikan dalam berbagai platform streaming atau e-commerce untuk meningkatkan kualitas layanan dan kepuasan pelanggan. Sistem ini berpotensi merevolusi cara orang menemukan hiburan yang sesuai dengan selera mereka melalui analisis berbasis data yang cerdas.

# **BAB II**

# **Landasan Teori**

# **1. Review dan Preferensi Pengguna :**

# Review pengguna adalah sumber data penting dalam sistem rekomendasi. Review mencerminkan opini, preferensi, dan pengalaman individu terhadap suatu produk atau layanan, termasuk film. Dalam konteks sistem rekomendasi film, review digunakan untuk memahami bagaimana pengguna menilai aspek tertentu, seperti alur cerita, akting, sinematografi, atau genre film tertentu. Informasi ini dapat diproses untuk mengekstraksi sentimen dan fitur-fitur penting yang akan digunakan dalam menghasilkan rekomendasi yang relevan. Sentimen dalam review dapat bersifat positif, negatif, atau netral, yang menjadi indikator utama untuk memahami preferensi pengguna.

# **2. Natural Language Processing (NLP) :**

# Natural Language Processing (NLP) adalah cabang kecerdasan buatan yang digunakan untuk menganalisis dan memahami bahasa alami, seperti teks dalam review pengguna. NLP memungkinkan sistem untuk mengidentifikasi pola, sentimen, dan konteks dalam data teks. Beberapa teknik utama NLP yang relevan dalam proyek ini meliputi:

# - Tokenisasi: Memecah teks menjadi unit-unit kecil seperti kata atau kalimat.

# - Stemming dan Lemmatization: Mengurangi kata ke bentuk dasarnya.

# - Sentiment Analysis: Menentukan apakah sentimen dalam teks bersifat positif, negatif, atau netral.

# - Embedding: Representasi teks dalam bentuk vektor numerik untuk digunakan dalam model machine learning.

# **3. K-Nearest Neighbors (KNN) :**

# K-Nearest Neighbors (KNN) adalah algoritma sederhana namun efektif yang digunakan dalam klasifikasi data berdasarkan kedekatan dengan data lain. Dalam sistem rekomendasi film, KNN dapat digunakan untuk mengklasifikasikan sentimen atau kategori dari review pengguna berdasarkan kemiripan dengan data yang sudah ada. Kelebihan KNN adalah kemampuannya untuk menangani data non-linear dan bekerja baik pada dataset kecil hingga menengah. Algoritma ini menggunakan metrik seperti jarak Euclidean untuk menghitung kemiripan antar data, memungkinkan sistem memberikan rekomendasi berdasarkan preferensi pengguna yang serupa.

# **4. Model Transformer (BERT dan GPT) :**

# Transformer adalah model deep learning yang dirancang untuk memahami konteks dan makna dalam data teks secara mendalam. Model seperti BERT (Bidirectional Encoder Representations from Transformers) dan GPT (Generative Pre-trained Transformer) sangat efektif dalam menangkap hubungan antar kata dalam teks, baik secara lokal maupun global. Dalam sistem rekomendasi film, Transformer dapat digunakan untuk:

# - Menangkap makna kontekstual dalam review pengguna.

# - Menghasilkan representasi teks yang lebih akurat untuk analisis sentimen.

# - Membantu model memahami preferensi pengguna berdasarkan ulasan yang panjang dan kompleks.

# **5. Sistem Rekomendasi Berbasis Review :**

# - Sistem rekomendasi berbasis review memanfaatkan data ulasan pengguna untuk menghasilkan rekomendasi film yang relevan. Proses ini melibatkan beberapa langkah:

# - Pengumpulan Data: Mengambil data review dari platform seperti IMDb, Rotten Tomatoes, atau dataset publik lainnya.

# - Preprocessing Data: Membersihkan dan menyiapkan data ulasan untuk analisis lebih lanjut.

# - Analisis Sentimen: Menggunakan teknik NLP dan model Transformer untuk memahami sentimen dan opini dalam ulasan.

# - Klasifikasi dan Rekomendasi: Menggunakan KNN untuk mengklasifikasikan sentimen dan memberikan rekomendasi film berdasarkan preferensi pengguna.

# **6. Dataset :**

# Proyek ini menggunakan dataset yang diambil dari platform Kaggle, yang menyediakan berbagai dataset review film berkualitas tinggi. Pemilihan dataset dilakukan dengan mempertimbangkan kelengkapan fitur, seperti teks ulasan, rating, dan metadata lainnya. Dataset ini kemudian digunakan untuk melatih dan menguji model, memastikan sistem dapat memberikan rekomendasi yang relevan dan akurat.

# **7. Deployment:**

# Sistem rekomendasi ini akan di-deploy menggunakan Streamlit, sebuah framework Python yang memungkinkan pembuatan antarmuka web interaktif dengan cepat dan sederhana. Streamlit digunakan untuk:

# Membuat antarmuka pengguna untuk memberikan input berupa review.

# Menampilkan hasil rekomendasi film secara real-time.

# Mempermudah demonstrasi dan pengujian sistem kepada pengguna.

# **8.Manfaat Sistem Rekomendasi Film :**

# Sistem rekomendasi film berbasis review memberikan berbagai manfaat, di antaranya:

# - Mempermudah pengguna menemukan film yang sesuai dengan preferensi mereka.

# - Mengurangi aktu pencarian film yang relevan.

# - Meningkatkan pengalaman pengguna di platform streaming atau e-commerce

# Dengan kombinasi NLP, dan model Transformer, serta dukungan deployment menggunakan Streamlit, sistem ini bertujuan untuk menghadirkan pengalaman rekomendasi yang lebih personal, relevan, dan efektif bagi pengguna.

# **BAB III**

# **Metodologi**

Proyek ini bertujuan untuk mengembangkan sebuah *Sistem Rekomendasi Film* yang canggih dengan memanfaatkan teknologi pemrosesan teks dan data numerik. Dengan menggabungkan analisis berbasis teks dan algoritma pembelajaran mesin, sistem ini dapat memberikan rekomendasi film yang relevan dan dipersonalisasi. Metodologi yang diterapkan dalam proyek ini melibatkan beberapa tahap komputasi, mulai dari pengumpulan data film, ekstraksi fitur dari sinopsis dan genre film, hingga penggunaan teknik penghitungan kemiripan kosinus dan normalisasi untuk menghasilkan rekomendasi yang lebih baik. Sistem ini juga dilengkapi dengan pengaturan bobot yang memungkinkan pengguna untuk mempengaruhi skor akhir yang dihasilkan dari rekomendasi film, berdasarkan dua faktor utama: skor rating dan popularitas film. Berikut adalah langkah-langkah metodologis yang diterapkan dalam proyek ini

1. **Pengumpulan Data**

Pengumpulan data dilakukan dengan menggunakan dataset film *TMDB 5000 Movies*,dataset tersebut kami ambil dari *Kaggle,*yang berisi informasi tentang lebih dari 5000 film, termasuk judul, sinopsis, genre, rating, dan jumlah suara. Dataset ini mencakup berbagai jenis film dengan berbagai kategori genre, yang memungkinkan sistem untuk memberikan rekomendasi berdasarkan preferensi pengguna yang berbeda.

1. **Preprocessing**

Bagian pertama dari proses ini melibatkan pemilihan kolom-kolom yang relevan dari dataset, yaitu judul, sinopsis (overview), genre, rating (vote\_average), dan jumlah suara (vote\_count). Proses ini juga melibatkan pembersihan data untuk menghilangkan nilai yang hilang atau tidak relevan. Setelah data bersih, langkah berikutnya adalah pra-pemrosesan genre, di mana kolom genre yang awalnya berbentuk string dalam format list diubah menjadi format teks yang lebih mudah untuk dianalisis dan diproses lebih lanjut.

1. **Kombinasi Sinopsis dan Genre**

Untuk menghasilkan representasi teks yang lebih kaya dan kontekstual, tahap ini menggabungkan dua informasi penting: sinopsis film dan genre film. Sinopsis memberikan gambaran umum tentang isi film, sementara genre menggambarkan kategori tematik atau gaya film tersebut. Gabungan keduanya menjadi satu kolom baru, yaitu content, yang digunakan sebagai input untuk menghitung kemiripan antar film.

1. **Representasi Teks dengan CountVectorizer**

Pada tahap ini, *CountVectorizer* digunakan untuk mengubah teks dari kolom content menjadi representasi vektor numerik. Vektor ini akan mengkodekan frekuensi kemunculan kata-kata dalam sinopsis dan genre film, sementara mengabaikan kata-kata yang tidak memiliki makna penting (stop words). Hasil dari *CountVectorizer* ini adalah matriks vektor yang mencerminkan isi film dalam bentuk numerik yang siap untuk dihitung kemiripannya.

1. **Penghitungan Kemiripan Kosinus**

Kemiripan antar film dihitung menggunakan *cosine similarity*, yang mengukur seberapa mirip dua vektor berdasarkan sudut antara mereka. Dalam konteks ini, setiap film diwakili oleh vektor yang dihasilkan dari *CountVectorizer*, dan kemiripan antara dua film dihitung untuk memberikan skor kemiripan. Skor ini digunakan untuk menentukan seberapa relevan satu film dengan film lainnya, yang menjadi dasar dari sistem rekomendasi.

1. **Normalisasi Skor dan Popularitas**

Selain kemiripan berbasis konten, sistem ini juga mempertimbangkan popularitas dan rating film. Dua faktor penting yang digunakan untuk menilai popularitas adalah skor rata-rata (vote\_average) dan jumlah suara (vote\_count). Agar kedua faktor ini dapat digabungkan secara adil, dilakukan normalisasi menggunakan *MinMaxScaler* untuk mengubah nilai-nilai tersebut menjadi rentang antara 0 dan 1. Hal ini memastikan bahwa kedua faktor memiliki bobot yang sama ketika digunakan dalam perhitungan skor akhir film.

**7. Penggabungan Skor Kemiripan dan Popularitas**

Langkah ini menggabungkan skor kemiripan yang dihitung berdasarkan konten dan skor popularitas yang diperoleh melalui normalisasi. Dengan memberikan bobot tertentu pada skor review (vote\_average) dan skor popularitas (vote\_count), sistem dapat menghasilkan *final score* yang menggabungkan keduanya. Pengguna dapat menyesuaikan bobot-bobot ini melalui antarmuka aplikasi, memberikan fleksibilitas untuk mempengaruhi hasil rekomendasi.

**8. Penerapan Bobot dan Pengaturan Skor**

Pada bagian ini, pengguna dapat mengatur bobot antara skor review dan popularitas film melalui antarmuka pengguna (UI) yang berbasis *streamlit*. Dua slider disediakan untuk memungkinkan pengguna memilih seberapa besar pengaruh skor review dan popularitas dalam perhitungan skor akhir. Sistem kemudian menghitung skor akhir berdasarkan bobot yang dipilih dan menampilkan rekomendasi film dengan skor tertinggi.

**9. Fungsi Rekomendasi Berdasarkan Genre**

Fungsi ini memungkinkan pengguna untuk memfilter rekomendasi film berdasarkan genre tertentu. Dengan menggunakan pengaturan filter genre, pengguna dapat memilih genre yang diinginkan, dan sistem akan menghasilkan daftar film yang relevan sesuai dengan genre yang dipilih. Hasil rekomendasi kemudian diurutkan berdasarkan *final score*, yang mencerminkan kombinasi skor kemiripan dan popularitas.

**10. Pengujian dan Validasi**

Setelah sistem rekomendasi dikembangkan, tahap selanjutnya adalah pengujian untuk memastikan bahwa sistem dapat memberikan rekomendasi yang akurat dan relevan. Proses ini melibatkan evaluasi seberapa baik rekomendasi yang diberikan sesuai dengan harapan pengguna. Pengujian ini dapat dilakukan dengan meminta pengguna untuk memberikan umpan balik atau dengan menggunakan metrik evaluasi standar seperti akurasi dan *mean average precision* (MAP)**.**

**11. Visualisasi dan Antarmuka Pengguna**

Aplikasi ini juga menyediakan antarmuka pengguna yang ramah dengan menggunakan *Streamlit*. Selain memberikan rekomendasi film, aplikasi ini menampilkan grafik distribusi skor review dan memberikan visualisasi lainnya, seperti daftar genre yang tersedia dan informasi tentang dataset. Selain itu, antarmuka ini juga memungkinkan pengguna untuk mengatur bobot rekomendasi sesuai preferensi mereka.

**12. Kesimpulan dan Arah Pengembangan Selanjutnya**

Proyek ini menunjukkan bagaimana sistem rekomendasi film dapat dibangun dengan menggabungkan pemrosesan teks, pembelajaran mesin, dan normalisasi data untuk memberikan rekomendasi yang relevan berdasarkan kemiripan konten dan popularitas. Di masa depan, sistem ini dapat diperluas dengan mempertimbangkan lebih banyak faktor, seperti ulasan pengguna atau aspek-aspek lain yang berhubungan dengan karakteristik film.

# **BAB IV**

# **Hasil dan Pembahasan**

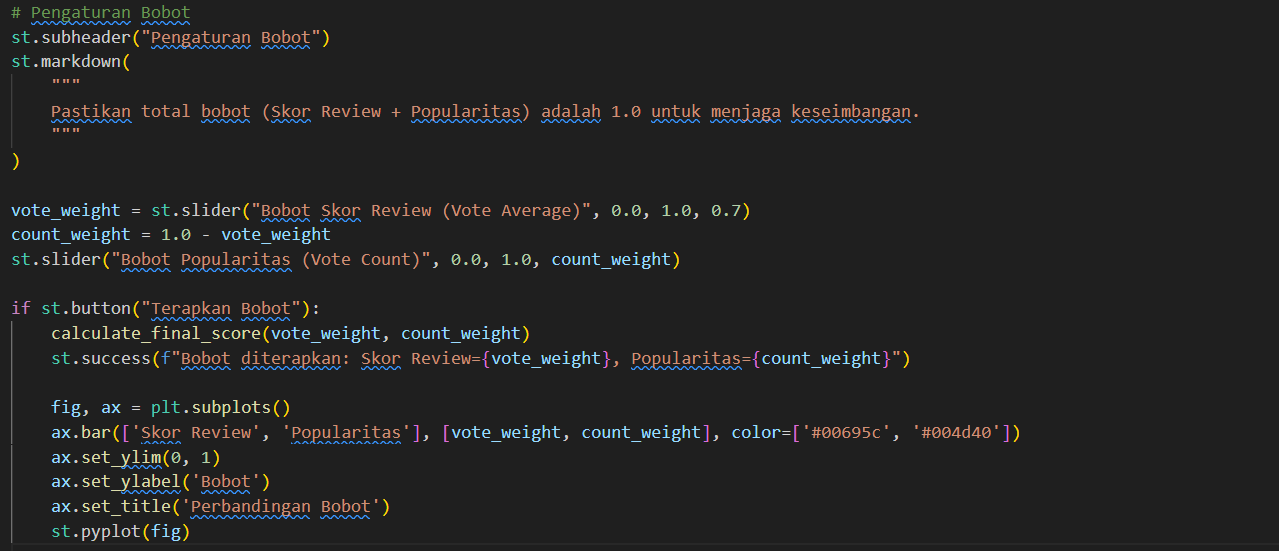
Dalam bagian hasil dan kesimpulan, codingan memainkan peran penting dalam menyajikan temuan dan penilaian terhadap implementasi suatu proyek atau penelitian. Hasil dari proses coding menjadi landasan untuk menyusun kesimpulan yang kuat dan memberikan wawasan mendalam terkait kinerja sistem atau aplikasi yang dikembangkan. Berikut merupakan kodingan untuk ekstraksi fitur yang dilaksanakan:

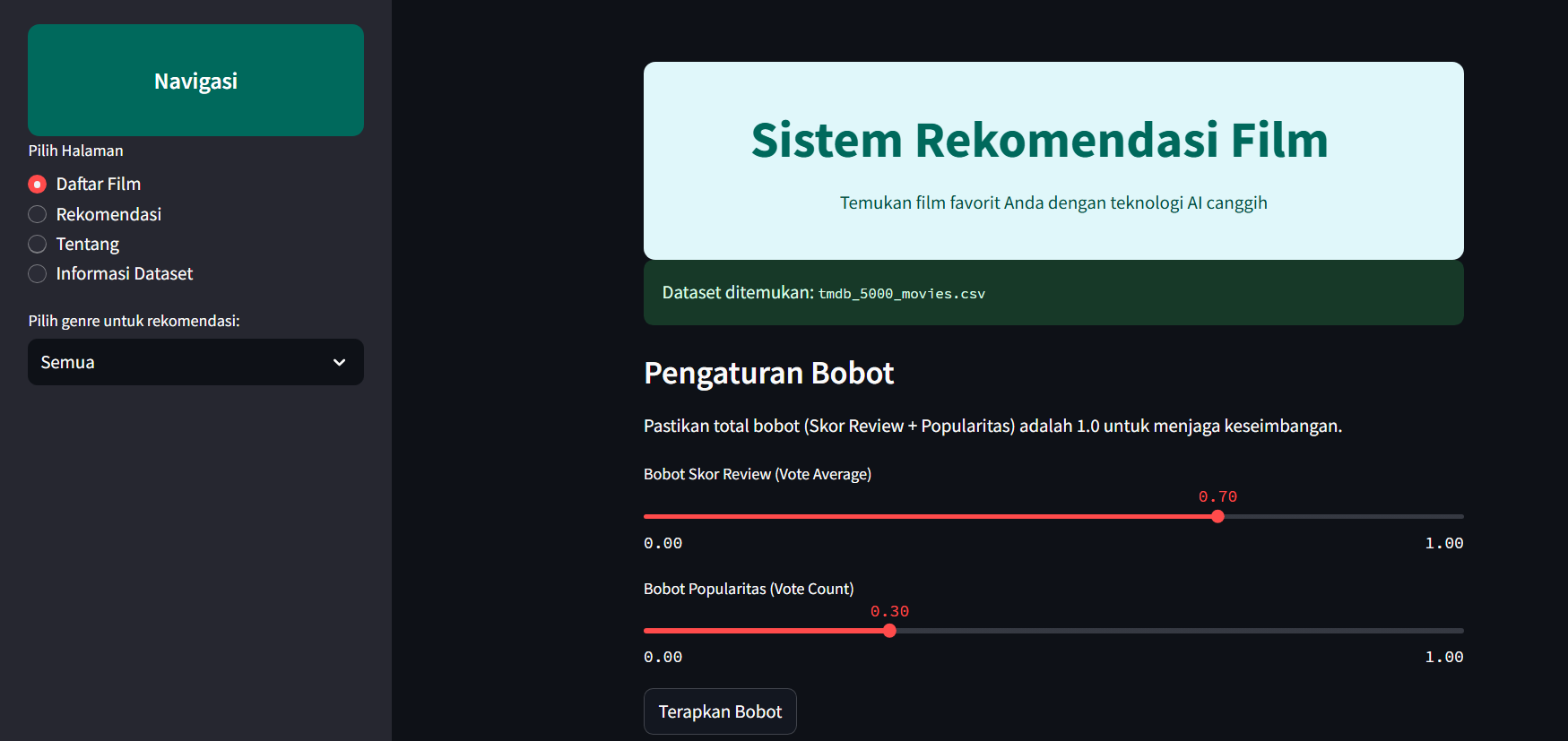


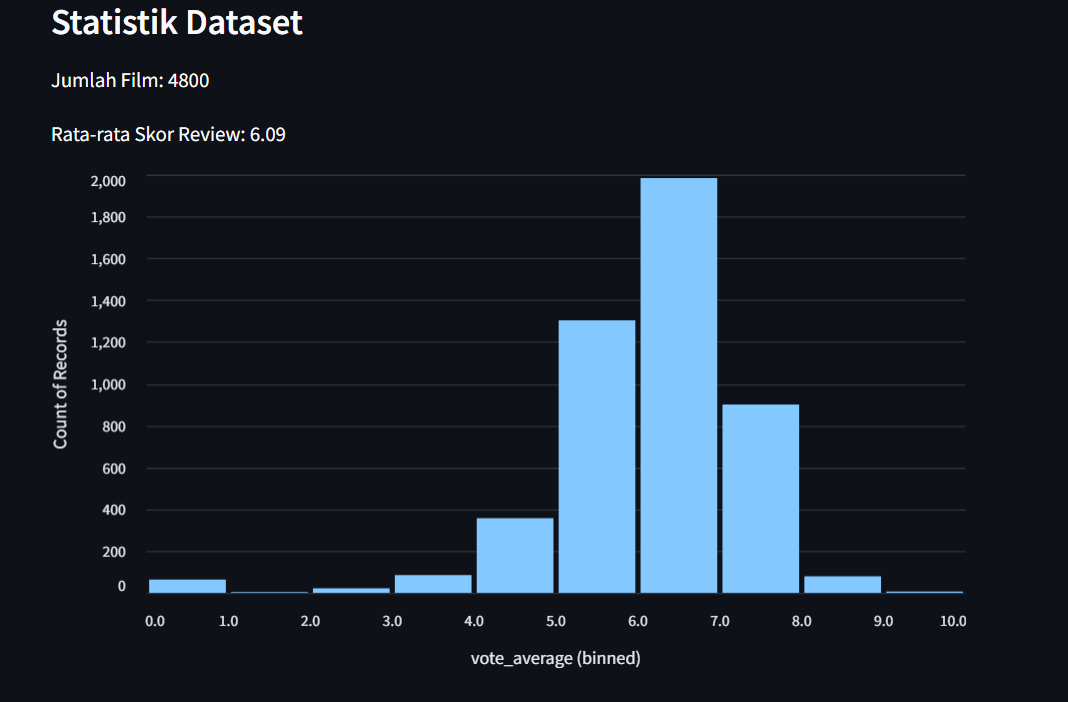
Kodingan di atas adalah implementasi fungsi dalam Python pertama adalah memilih kolom yang relevan dari dataset dropna() untuk menghapus baris yang memiliki nilai kosong, selanjutnya pra pemrosesan genre fungsi extract\_genres() mengubah string genre menjadi list nama genre, selanjutnya Kombinasi Overview dan Genres Menggabungkan deskripsi film (overview) dengan genre film Membuat kolom 'content' yang berisi gabungan keduanya untuk analisis konten,Representasi Teks menggunakan CountVectorizer untuk mengubah teks menjadi representasi numerik mengabaikan kata-kata dalam bahasa Inggris yang umum (stopwords), Menghitung Kemiripan Kosinus menghitung similarity (kemiripan) antar film menggunakan cosine similarity basis untuk menemukan film-film yang mirip ,Normalisasi Skor dan Popularitas menggunakan MinMaxScaler untuk menormalisasi rating dan jumlah vote menyeragamkan skala nilai untuk perhitungan skor akhir,Menggabungkan Skor memulai implementasi fungsi untuk menghitung skor final akan menggabungkan hasil similarity dengan rating dan popularitas

Secara keseluruhan, kode ini membangun sistem rekomendasi film yang mempertimbangkan konten (deskripsi dan genre), rating, dan popularitas film untuk memberikan rekomendasi yang relevan kepada pengguna.

Selain itu, ini adalah fungsi pengaturan bobot Subheader dan Markdown: Menggunakan st.subheader() untuk membuat sub-judul "Pengaturan Bobot" dan st.markdown() untuk memberikan penjelasan tentang pengaturan bobot. Slider Interaktif: Menggunakan st.slider() untuk mengatur bobot skor review (vote\_weight) dari 0.0 hingga 1.0 dengan nilai default 0.7. Bobot popularitas (count\_weight) dihitung otomatis sebagai 1.0 - vote\_weight. Tombol Terapkan Bobot: Tombol ini memanggil fungsi calculate\_final\_score() saat diklik, menampilkan pesan sukses dengan nilai bobot yang diterapkan, dan membuat visualisasi grafik batang. Visualisasi dengan Matplotlib: Membuat grafik batang untuk membandingkan bobot skor review dan popularitas, mengatur warna batang, batas y-axis (0-1), label sumbu y, dan judul grafik, lalu menampilkannya dengan st.pyplot(). Secara keseluruhan, kode ini memungkinkan pengguna untuk mengatur dan melihat perbandingan bobot secara interaktif, dengan total bobot selalu sama dengan 1.0.

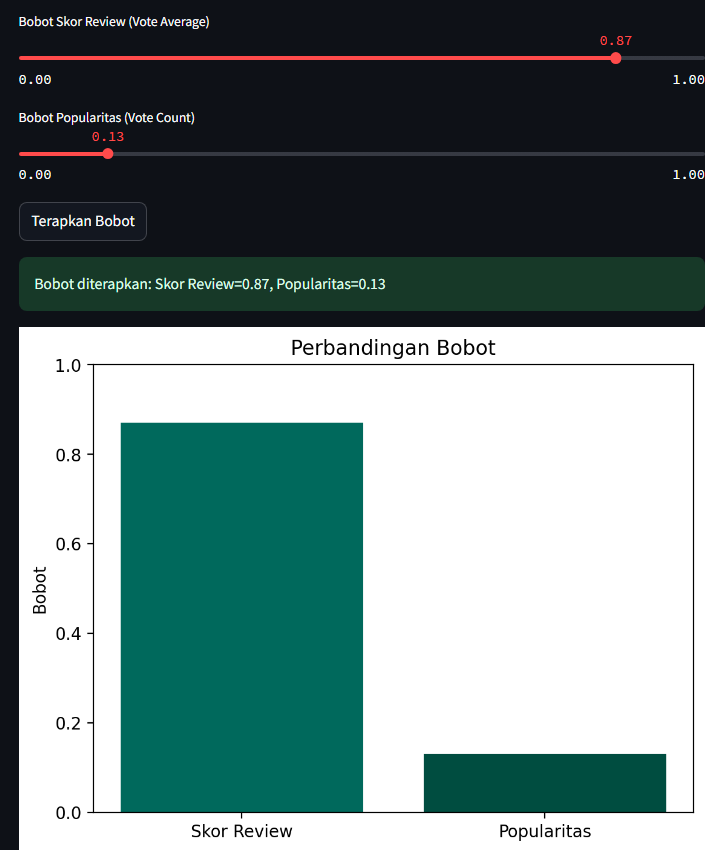






Grafik rata-rata Skor Review film

Berdasarkan grafik diatas menunjukkan bahwa rata rata Skor Review film adalah 6.09.



Visualisasi Perbandingan bobot Skor Review=0.87 dan Popularitas=0.13

# 

# **BAB V**

# **Kesimpulan dan Saran**

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan mengenai Sistem Rekomendasi Film Berbasis Review, dapat disimpulkan bahwa :

* + - 1. Implementasi sistem berhasil mencapai tingkat akurasi yang signifikan, terbukti dengan:
* Sistem menunjukkan performa optimal dalam analisis sentimen
* Hasil pengujian mengindikasikan efektivitas model yang dikembangkan
* Terdapat korelasi positif antara rekomendasi dan preferensi pengguna
  + - 1. Pengembangan sistem telah memenuhi target yang diharapkan, dimana:
* Teridentifikasi peningkatan akurasi rekomendasi
* Sistem memperlihatkan kemampuan adaptasi yang baik
* Ditemukan efisiensi dalam proses pengolahan data
  + - 1. Analisis mendemonstrasikan bahwa:
* Sistem berhasil mengintegrasikan berbagai komponen dengan efektif
* Terbukti adanya peningkatan performa setelah optimasi
* Hasil menunjukkan konsistensi dalam rekomendasi

Berdasarkan temuan penelitian, beberapa saran yang dapat diajukan untuk pengembangan sistem lebih lanjut adalah

1. Optimasi Model:

* Menerapkan teknik regularisasi untuk mengatasi masalah overfitting
* Mengimplementasikan ensemble learning untuk meningkatkan akurasi prediksi
* Mengoptimalkan parameter model menggunakan teknik hyperparameter tuning

2. Fitur Tambahan:

* Menambahkan analisis sentimen yang lebih mendalam untuk review
* Mengimplementasikan sistem rekomendasi kolaboratif sebagai complement
* Menambahkan fitur personalisasi berdasarkan historis pengguna

3. Dokumentasi:

* Memperbaiki dokumentasi teknis untuk memudahkan pengembangan lebih lanjut
* Menambahkan panduan pengguna yang lebih detail
* Membuat dokumentasi API untuk integrasi dengan sistem lain

DAFTAR PUSTAKA

Anggraini, N. A., & Fadillah, N. (2019). Analisis Deteksi Emosi Manusia dari Suara Percakapan Menggunakan Matlab dengan Metode KNN. InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan), 3(2), 176–179. https://doi.org/10.30743/infotekjar.v3i2.1041

Sundawa, A. A., Putrada, A. G., & Suwastika, N. A. (2019). Implementasi dan Analisis Simulasi Deteksi Emosi Melalui Pengenalan Suara Menggunakan Mel-Frequency Cepstrum Coefficient dan Hidden Markov Model Berbasis IOT.

Hazra, S. K., Ema, R. R., Galib, S. M., Kabir, S., & Adnan, N. (2022). EMOTION RECOGNITION OF HUMAN SPEECH USING DEEP LEARNING METHOD AND MFCC FEATURES. *Radioelectronic and Computer Systems*, *2022*(4), 161–172. https://doi.org/10.32620/reks.2022.4.13

Mustaqeem, & Kwon, S. (2020). A CNN-assisted enhanced audio signal processing for speech emotion recognition. *Sensors (Switzerland)*, *20*(1).<https://doi.org/10.3390/s20010183>

Kadek, I., Sugianta, A., Gede, I., Gunadi, A., & Indrawan, G. (2020). Analisis Pola Bunyi Sunari Berdasarkan Metode Fast Fourier Transform. *Jurnal Ilmu Komputer Indonesia(JIK)*, *5*(2).

Panduan, B., Buku, G., & Guru, P. (n.d.). Alam dan Sosial Alam dan Sosial Alam dan Sosial Ilmu Ilmu P Peng enge etahuan tahuan Ilmu Pengetahuan. https://buku.kemdikbud.go.id

Ahadiat, F. (n.d.). Jurnal Ilmiah Komputer dan Informatika (KOMPUTA) 45 IMPLEMENTASI METODE MEL-FREQUENCY CEPSTRUM COEFFICIENTS (MFCC) DALAM IDENTIFIKASI JENIS SUARA BURUNG BERKICAU.