**Analisis dan Rekomendasi Film Netflix Berdasarkan Genre, Tahun, dan Negara**



Disusun oleh :

Esadhipa Raif S - A11.2022.14233

**UNIVERSITAS DIAN NUSWANTORO**

**FAKULTAS ILMU KOMPUTER**

**TEKNIK INFORMATIKA**

**2024**

**Ringkasan:**

Sistem yang dikembangkan dalam proyek ini adalah rekomendasi film berbasis genre, tahun, dan negara. Sistem ini memanfaatkan dataset Netflix dan pendekatan berbasis TF-IDF untuk menganalisis teks dan menghitung kesamaan antar genre, yang diintegrasikan dengan fungsi pencarian filter untuk memberikan hasil yang sesuai kriteria pengguna.

**Permasalahan:**

Netflix menyediakan berbagai film dan serial TV dengan genre, tahun rilis, dan wilayah yang beragam. Dengan banyaknya pilihan, pengguna membutuhkan alat untuk merekomendasikan konten yang relevan.

**Tujuan:**

1. Mengembangkan sistem yang merekomendasikan film berdasarkan kriteria pengguna.
2. Meningkatkan pengalaman pengguna dengan memberikan rekomendasi berkualitas tinggi.

**Model:**

Model yang digunakan dalam proyek ini adalah *TF-IDF (Term Frequency-Inverse Document Frequency),* yang merupakan teknik untuk merepresentasikan teks sebagai vektor numerik, dan *cosine similarity*, yang digunakan untuk mengukur tingkat kemiripan antara genre pilihan pengguna dengan data dalam dataset.

**Dataset:**

1. **Sumber**: Dataset Netflix di ambil dari kaggle
2. **Kolom Penting**:

● title: Judul konten.

● genres: Genre film.

● releaseYear: Tahun rilis.

● imdbAverageRating: Rating IMDb.

● availableCountries: Negara tempat film tersedia.

**Exploratory Data Analysis (EDA)**:

* Distribusi genre menunjukkan dominasi genre "Drama" dan "Komedi".
* Distribusi tahun rilis memuncak pada tahun 2000-an.

**Pembersihan Data**:

* Mengisi nilai kosong pada kolom seperti imdbAverageRating dengan rata-rata.
* Mengganti nilai kosong di genres dengan "Other".

**Feature Engineering**:

* **TF-IDF**: Digunakan untuk merepresentasikan kolom genre sebagai vektor numerik.
* **Cosine Similarity**: Digunakan untuk mencocokkan genre pencarian dengan dataset.

**Proses Learning / Modeling**

**Pendekatan Model**:

1. **Search Function**:

● Input: Tipe konten, genre, tahun, negara.

● Filter data sesuai kriteria.

● Urutkan berdasarkan rating IMDb.

1. **Rekomendasi Berbasis TF-IDF**:

● Representasi teks genre dalam vektor numerik.

● Gunakan cosine similarity untuk menentukan kecocokan genre.

**Performa Model**

**Search Function**:

● Mampu memfilter dan mengurutkan film sesuai kriteria pengguna.

● Contoh output: Daftar film dengan genre "Horror" di "Indonesia".

**TF-IDF Recommendation**:

Memberikan skor kemiripan berbasis genre.

● Output: Film dengan skor kemiripan tertinggi untuk permintaan pencarian tertentu.

**Hasil**:

1. Sistem berhasil memberikan rekomendasi film yang relevan.
2. TF-IDF efektif dalam mencocokkan genre berdasarkan permintaan pengguna.

**Kesimpulan**:

1. Metode berbasis pencarian dan TF-IDF mampu meningkatkan pengalaman pengguna dalam menemukan konten yang sesuai.
2. Sistem ini dapat diadaptasi untuk platform lain dengan dataset yang serupa.