

Tugas Modul 2

Akuisisi Data: Fondasi Proyek Data di Industri

Nama Dataset: Netflix TV Shows and Movies – *titles.csv*

Sumber: Kaggle — “[Netflix TV Shows and Movies](#)” oleh Victor Soeiro

Data set ini berisi daftar semua acara yang tersedia di Netflix streaming yang tersedia pada Juli 2022 dan berisi data yang tersedia di wilayah Amerika Serikat. Data set ini berisi ±5.850 judul unik di Netflix dengan 15 kolom (file *titles.csv*) yang berisi informasi, termasuk:

No	Nama Kolom	Keterangan
1.	id	ID judul di JustWatch
2.	title	Nama film atau TV show
3.	show_type	Apakah ini “MOVIE” atau “SHOW”
4.	description	Deskripsi singkat dari konten
5.	release_year	Tahun rilis
6.	age_certification	Rating usia (mis. PG, TV-MA, R, dll.)
7.	runtime	Durasi dalam menit (film atau durasi episode)
8.	genres	Genre / jenis konten (bisa banyak genre per judul)
9.	production_countries	Negara yang memproduksi konten tersebut
10.	seasons	Jumlah musim (untuk acara TV)
11.	imdb_id	ID konten di IMDb
12.	imdb_score	Nilai konten berdasarkan penilaian pengguna IMDb
13.	imdb_votes	Jumlah suara (votes) di IMDb
14.	tmdb_popularity	Popularitas di situs TMDB
15.	tmdb_score	Nilai / skor di TMDB

Pertanyaan Replikasi:

1. Berapa jumlah film dibandingkan TV show di dataset ini?
2. Tahun berapa paling banyak film atau TV show dirilis?
3. Negara mana yang paling banyak memproduksi film/show?
4. Genre apa yang paling sering muncul dalam dataset ini?
5. Berapa rata-rata skor IMDb dari seluruh film?

Pertanyaan Baru:

1. Apakah film dengan rating usia R/TV-MA memiliki skor IMDb lebih tinggi dibandingkan rating usia lain?
2. Apakah durasi film berhubungan dengan skor IMDb (misalnya, film yang lebih panjang cenderung lebih tinggi nilainya)?
3. Apakah popularitas film/TV show di TMDB berhubungan dengan skor IMDb?
4. Negara mana yang paling banyak menghasilkan film dengan IMDb score di atas 8?
5. Bagaimana tren kualitas film dari tahun 1950–2020 jika dilihat dari rata-rata skor IMDb?