

**EKSPLORASI DAN VISUALISASI DATA KONTEN NETFLIX
BERDASARKAN PERBANDINGAN FILM DAN ACARA
TV YANG DIBUAT OLEH NEGARA DI DUNIA**



Disusun Oleh:

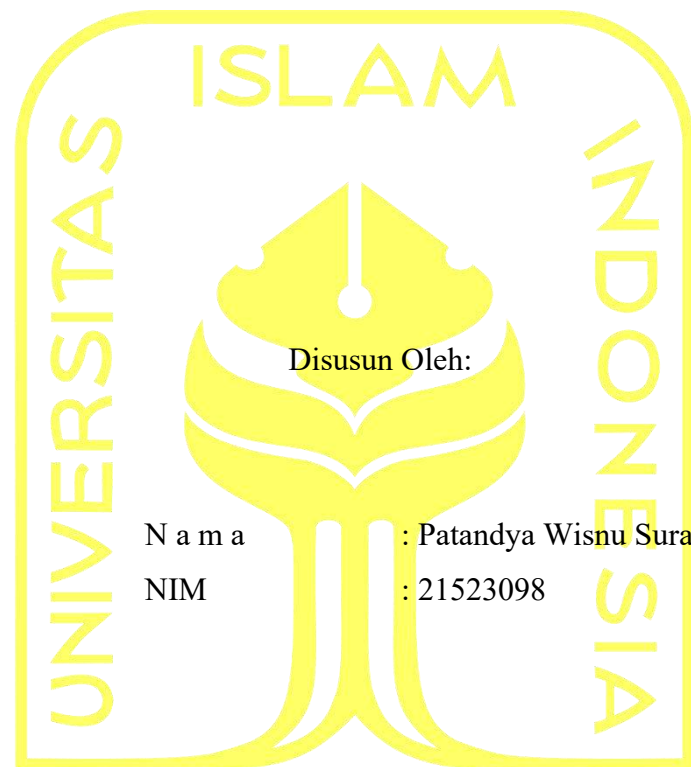
N a m a : Patandya Wisnu Suraya
NIM : 21523098

**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
2025**

HALAMAN PENGESAHAN DOSEN PEMBIMBING

**EKSPLORASI DAN VISUALISASI DATA KONTEN NETFLIX
BERDASARKAN PERBANDINGAN FILM DAN ACARA
TV YANG DIBUAT OLEH NEGARA DI DUNIA**

TUGAS AKHIR



N a m a : Patandya Wisnu Suraya
NIM : 21523098

الجامعة الإسلامية
الاندونيسية

Yogyakarta, 09 Mei 2025

Pembimbing,


SRI MULYATI, S.KOM., M.KOM

HALAMAN PENGESAHAN DOSEN PENGUJI

**EKSPLORASI DAN VISUALISASI DATA KONTEN NETFLIX
BERDASARKAN PERBANDINGAN FILM DAN ACARA
TV YANG DIBUAT OLEH NEGARA DI DUNIA**

TUGAS AKHIR

Telah dipertahankan di depan sidang pengujian sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 26 Mei 2025

Tim Penguji

Sri Mulyati, S.Kom., M.Kom.

Anggota 1

Arrie Kurniawardhani, S.Si., M.Kom.

Anggota 2

Dr. Feri Wijayanto, S.T., M.T.



Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



(Thomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Patandya Wisnu Suraya

NIM : 21523098

Tugas akhir dengan judul:

EKSPLORASI DAN VISUALISASI DATA KONTEN NETFLIX BERDASARKAN PERBANDINGAN FILM DAN ACARA TV YANG DIBUAT OLEH NEGARA DI DUNIA

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 29 Juni 2024



A handwritten signature in black ink, appearing to read "Patandya Wisnu Suraya".

Patandya Wisnu Suraya

HALAMAN PERSEMBAHAN

Tugas akhir ini saya persembahkan sebagai bentuk dedikasi dan kontribusi saya bagi siapa saja yang ingin mempelajari pengolahan data lebih dalam. Semoga karya ini dapat menjadi sumber wawasan dan inspirasi dalam memahami dunia data yang semakin berkembang.

HALAMAN MOTO

“Dan katakanlah, “Kebenaran telah datang dan yang batil telah lenyap.” Sesungguhnya yang batil itu pasti lenyap. ”

- Al-Isra: 81

“Bersabarlah kamu dan kuatkanlah kesabaranmu dan tetaplah bersiaga-siaga dan bertaqwalah kepada Allah supaya kamu menang.”

- Ali-Imran: 200

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh.

Alhamdulillah rabbil alamin, Segala puji dan syukur penulis panjatkan kepada Allah SWT Tuhan semesta alam, tuhan yang mengetahui, tuhan yang maha baik, yang selalu melampiahkan karunia-Nya sehingga penulis dapat dengan lancar menyelesaikan tugas akhir ini. Shalawat serta salam kepada Nabi Muhammad SAW yang telah membawa umat islam dari jaman jahilliyah kepada jaman yang terang benderang.

Setiap tahap penyusunan ini, tentunya penulis tidak lepas dari arahan serta dukungan dari berbagai pihak. Penulis mengucapkan terima kasih serta rasa hormat sebesar-besarnya kepada semua pihak yang telah membantu. Dengan segala hormat, ijinkan penulis untuk mengucapkan terima kasih kepada:

1. Allah SWT, Tuhan semesta alam yang selalu memberikan rahmat serta karuniannya kepada penulis di segala lini kehidupan dari penulis lahir hingga wafat kelak.
2. Orang tua yang selalu memberikan dukungan, doa, dan bantuan yang sangat berharga baik secara moril maupun materiil, hingga penulis berhasil menamatkan skripsi.
3. Kakak dan Adik penulis yang senantiasa memberikan dukungan, semangat, dan motivasi yang tak ternilai.
4. Prof. Fathul Wahid, S.T., M.Sc., Ph.D., selaku Rektor Universitas Islam Indonesia.
5. Bapak Dr. Raden Teduh Dirgahayu, S.T., M.Sc., selaku Ketua Jurusan Informatika Universitas Islam Indonesia
6. Bapak DThomas Hatta Fudholi, S.T., M.Eng., Ph.D., selaku Ketua Program Studi Informatika Program Sarjana Fakultas Teknologi Industri Universitas Islam Indonesia
7. Ibu Sri Mulyati, S.Kom., M.Kom., selaku dosen pembimbing tugas akhir.
8. Bapak dan Ibu dosen Program Studi Informatika, yang telah berjasa dalam memberikan masukan serta ilmu yang bermanfaat selama penelitian ini.
9. Grup kopmer/penuh berkah melimpah, sahabat yang telah menemani penulis dari awal perkuliahan sampai hingga akhir, Vahri, Octa, Fikri, Fadhlhan, Ancol, Kipli, Bima, Agung, Ahsan, Maul, Wildan, Wawan, Luqman.
10. Teman-teman organisasi serta event yang telah membuat penulis agar bisa lebih berkembang dan mendapatkan pengetahuan.
11. Serta banyak pihak lainnya yang tidak bisa disebutkan namanya.

Penulis menyadari bahwa tugas akhir ini masih jauh dari sempurna. Oleh karena itu, Penulis sangat mengharapkan kritik, saran, dan masukan yang membangun dari pembaca agar penulisan ilmiah ini dapat lebih baik lagi di masa yang akan datang.

Akhir kata, penulis berharap skripsi ini dapat memberikan manfaat dan kontribusi yang positif bagi perkembangan ilmu pengetahuan di bidang analisis data pun yang berkaitan dengannya. Semoga tugas akhir ini dapat membuka wawasan dan memberikan inspirasi kepada pembaca yang mengkaji topik serupa di masa mendatang. Penulis mengakhiri kata pengantar ini dengan memohon maaf jika terdapat kekurangan dan kesalahan di dalamnya. Semoga penulisan ini dapat menjadi pijakan untuk penelitian lebih lanjut di masa yang akan datang.

Wassalamu'alaikum Wr. Wb.

Yogyakarta, 29 Juni 2024



Patandya Wisnu Suraya

SARI

Teknologi telah mengubah hampir semua aspek tata kelola bisnis, termasuk efisiensi operasional, pengambilan keputusan, hingga inovasi produk dan layanan. Salah satu contoh nyata adalah industri hiburan, khususnya platform layanan streaming seperti Netflix, yang memanfaatkan teknologi dan analisis data untuk mengembangkan serta merekomendasikan konten kepada pengguna.

Penelitian ini bertujuan untuk menganalisis industri hiburan pada platform Netflix berdasarkan konten yang diproduksi oleh berbagai negara di dunia. Data yang digunakan dalam penelitian ini berjumlah 8.807 baris dengan 12 atribut berbeda, yang mencakup informasi seperti jenis tayangan, negara asal, tanggal rilis, durasi, hingga klasifikasi usia. Analisis dilakukan dengan menerapkan metode *Exploratory Data Analysis (EDA)*, visualisasi data, serta algoritma K-means *Clustering*.

Hasil analisis menunjukkan bahwa jenis tayangan di Netflix didominasi oleh film (69,7% atau 6.126 entri), sementara acara TV mencakup 30,3% (2.664 entri). Dari segi waktu rilis, film memiliki rata-rata tahun rilis mundur sebesar 5,73 tahun dengan rentang maksimum 75 tahun, sedangkan acara TV rata-rata mundur selama 2,30 tahun dengan maksimum 93 tahun. Puncak perilisan film terjadi pada bulan Juli (565 rilis), sedangkan acara TV lebih sering dirilis pada bulan Desember (265 rilis).

Negara-negara seperti Amerika Serikat, India, dan Inggris tercatat sebagai produsen konten terbanyak, sementara konten dari negara-negara Asia dan Eropa juga menunjukkan potensi besar dalam menarik perhatian pengguna. Genre yang paling dominan adalah "*International Movies*" dengan 2.752 rilis, diikuti oleh genre "*Dramas*". Selain itu, konten anak-anak, khususnya dengan klasifikasi usia TV-Y dan TV-Y7, banyak ditemukan dalam perilisan acara TV, menunjukkan upaya Netflix dalam menyediakan tayangan ramah anak. Pada tahap klusterisasi, proses K-Means dilakukan secara terpisah untuk film dan TV Show. Untuk film, terbentuk dua kluster utama: kluster pertama berisi film-film lama dengan variasi durasi yang lebih luas, sedangkan kluster kedua didominasi oleh film-film modern dengan durasi yang lebih konsisten di sekitar 90–120 menit. Pada TV Show, kluster pertama berisi sebagian besar tayangan berdurasi pendek (1–3 episode), umumnya berupa mini-series modern, sedangkan kluster lainnya berisi konten dengan jumlah episode lebih banyak dan tahun rilis yang lebih bervariasi.

Kata kunci: Netflix, data analisis, *Exploratory Data Analysis (EDA)*, visualisasi data, K-means, Python.

GLOSARIUM

<i>Age ratings</i>	Klasifikasi usia untuk menilai kesesuaian konten berdasarkan kelompok umur tertentu.
Data Analisis	Proses mengumpulkan, mengolah, dan menganalisis data untuk mendapatkan wawasan yang berguna dalam pengambilan keputusan strategis.
EDA	Metode eksplorasi data untuk menemukan pola, anomali, dan hubungan dalam dataset melalui visualisasi dan statistik.
Genre	Kategori atau jenis konten yang mengelompokkan film dan acara TV berdasarkan tema atau gaya cerita.
K-means	Algoritma <i>Clustering</i> yang mengelompokkan data ke dalam sejumlah cluster berdasarkan kesamaan karakteristik.
Netflix	Platform layanan streaming digital yang menyediakan konten berupa film dan acara TV dari berbagai negara.
Tren	Pola atau kecenderungan yang menunjukkan arah perubahan dalam perilaku dan preferensi konsumen.
Visualisasi Data	Teknik penyajian data dalam bentuk grafik, diagram, atau chart untuk mempermudah interpretasi dan pengambilan keputusan.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING.....	ii
HALAMAN PENGESAHAN DOSEN PENGUJI	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO	vi
KATA PENGANTAR	vii
SARI.....	ix
GLOSARIUM.....	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Metodologi Penelitian	4
BAB II KAJIAN PUSTAKA	5
2.1 Penelitian Terdahulu	5
2.2 Data Analisis	8
2.3 Exploratory Data Analysis (EDA)	9
2.4 K-means	9
2.5 Elbow Method.....	9
2.6 Industri Hiburan	10
BAB III METODE PENELITIAN	11
3.1 <i>Data Requirement</i>	13
3.2 <i>Data Collection</i>	13
3.3 <i>Data Processing</i>	13
3.4 <i>Data Analysis</i>	14
3.5 <i>Visualize and interpret the result</i>	15

BAB IV HASIL DAN PEMBAHASAN	17
4.1 <i>Data Collection</i>	17
4.2 <i>Data Processing</i>	19
4.3 <i>Data Analysis</i>	24
4.4 <i>Visualize and interpret the result</i>	35
BAB V KESIMPULAN DAN SARAN.....	64
5.1 Kesimpulan	64
5.2 Saran.....	65
DAFTAR PUSTAKA	66
LAMPIRAN	68

DAFTAR TABEL

Tabel 4.1 Tabel rincian kolom data	18
--	----

DAFTAR GAMBAR

Gambar 1.1 Jumlah User Berlangganan Di Platform Netflix (Statista, 2022).....	2
Gambar 1.2 Metode data analisis yang akan digunakan.....	4
Gambar 3.1 metodologi penelitian yang digunakan dalam proses ini.....	12
Gambar 3.2 Langkah metode analisis data.....	12
Gambar 4.1 Halaman Netflix Dataset.....	17
Gambar 4.2 Netflix_titles.csv.....	17
Gambar 4.3 Data Tidak Memiliki Nilai.....	19
Gambar 4.4 Format Penanggalan Belum Sesuai.....	20
Gambar 4.5 Kode Pengecekan Duplikasi Data.....	20
Gambar 4.6 Kode Pembersihan Data.....	20
Gambar 4.7 Hasil penyesuain format data.....	22
Gambar 4.8 Boxplot.....	22
Gambar 4.9 Boxplot yang telah dilakukan IQR.....	23
Gambar 4.10 Kode pengecekan jumlah tv dan film.....	24
Gambar 4.11 Kode pemisahan genre pada konten.....	24
Gambar 4.12 Hasil pemisahan genre pada data film dan acara tv.....	25
Gambar 4.13 Kode untuk mengetahui durasi waktu rilis.....	25
Gambar 4.14 Hasil waktu yang digunakan untuk merilis film dan acara tv pada data.....	25
Gambar 4.15 Hasil waktu yang digunakan untuk merilis film dan acara tv pada data.....	26
Gambar 4.16 Hasil waktu yang digunakan untuk merilis masing- masing.....	26
Gambar 4.17 kode untuk menghitung konten per bulan.....	27
Gambar 4.18 Jumlah perilsan film dan acara tv pada setiap bulan.....	27
Gambar 4.19 Visualisasi Hasil Jumlah perilsan film dan acara tv pada setiap bulan.....	27
Gambar 4.20 kode untuk mengubah bulan ke 4 musim.....	28
Gambar 4.21 Hasil perilsan sesuai musim.....	28
Gambar 4.22 Hasil visualisasi perilsan sesuai musim.....	29
Gambar 4.23 kode untuk menemukan aktivitas konten yang sering dirilis pada suatu bulan.....	30
Gambar 4.24 Bulan paling dominan rilisnya film/acara tv.....	30
Gambar 4.25 Hasil visualisasi bulan juli sebagai dominan rilisnya film dan acara tv.....	30
Gambar 4.26 kode untuk menemukan jumlah klasifikasi usia pada konten.....	31
Gambar 4.27 Jumlah klasifikasi usia film/acara tv sesuai usia penonton.....	31

Gambar 4.28 Hasil Visualisasi jumlah klasifikasi usia film dan acara tv sesuai usia penonton	32
Gambar 4.29 kode untuk menemukan jumlah klasifikasi usia konten pada bulan	32
Gambar 4.30 Perilisan terbanyak klasifikasi usia film/acara tv pada bulan.	33
Gambar 4.31 Hasil Visualisasi perilisan terbanyak klasifikasi usia film dan acara tv pada bulan.	33
Gambar 4.32 Kode untuk menemukan total jumlah konten pada suatu negara.....	34
Gambar 4.33 Jumlah negara dalam perilisan film/acara tv.....	34
Gambar 4.34 Jumlah negara dalam perilisan film/acara tv.....	35
Gambar 4.35 Visualisasi perbandingan film dan acara tv	36
Gambar 4.36 Visualisasi perbandingan film dan acara tv per tahun	36
Gambar 4.37 Visualisasi top 5 tahun rilis konten per tahun	37
Gambar 4.38 Visualisasi genre film dan acara tv Netflix pada data.....	38
Gambar 4.39 Visualisasi perbandingan konten klasifikasi usia tertentu	39
Gambar 4.40 Visualisasi Distribusi perbandingan konten klasifikasi usia.....	41
Gambar 4.41 Visualisasi Top 10 Directors.....	42
Gambar 4.42 Visualisasi Top 10 Cast Members	43
Gambar 4.43 Visualisasi Top 10 Durasi TV Shows	43
Gambar 4.44 Visualisasi Top 10 Durasi Movies	44
Gambar 4.45 Visualisasi Top 10 Genres TV Shows	45
Gambar 4.46 Visualisasi Top 10 Genres Movies	46
Gambar 4.47 Visualisasi Top 3 Genres per Year	47
Gambar 4.48 Visualisasi Top 3 TV Show Genres per Year	48
Gambar 4.49 Visualisasi Top 3 Age Ratings per Year.....	49
Gambar 4.50 Visualisasi Top 3 Age Ratings Movies per Year.....	50
Gambar 4.51 Visualisasi Top 3 Age Ratings TV Shows per Year.....	51
Gambar 4.52 Visualisasi Top 3 Genres per Month.....	52
Gambar 4.53 Visualisasi Top 3 Movies Countries per Year	53
Gambar 4.54 Visualisasi Top 3 TV Shows Countries per Year	54
Gambar 4.55 Visualisasi Top 3 Countries per Year	55
Gambar 4.56 Visualisasi <i>elbow method</i> movies	56
Gambar 4.57 Visualisasi <i>silhouette</i> analisis movies	57
Gambar 4.58 Visualisasi K-means Movies.....	58
Gambar 4.59 Distribusi Film per Klaster Movies.....	59

Gambar 4.60 Visualisasi <i>elbow method</i> TV Shows.....	60
Gambar 4.61 Visualisasi <i>silhouette</i> analisis TV Shows.....	61
Gambar 4.62 Visualisasi K-means TV Shows.....	62
Gambar 4.63 Pembagian visualisasi sesuai kluster TV Shows.....	63

BAB I PENDAHULUAN

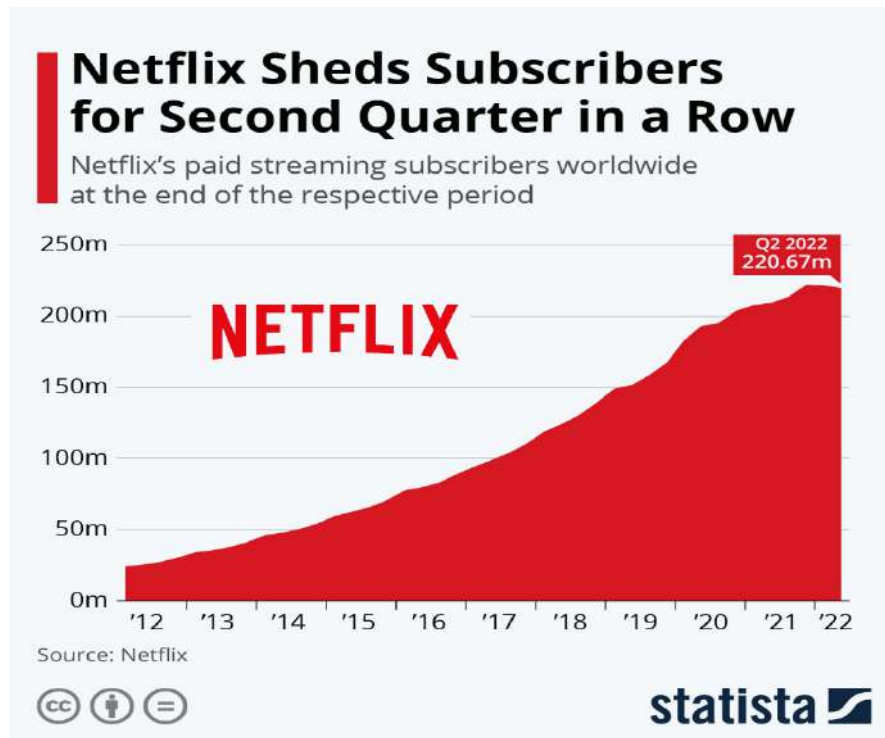
1.1 Latar Belakang

Dalam era digital yang terus berkembang pesat, menggunakan teknologi menjadi salah satu keharusan dalam menjalankan bisnis di berbagai sektor industri. Tidak dapat dimungkiri bahwa teknologi telah mengubah hampir semua aspek tata kelola bisnis yang berdampak kepada efisiensi operasional, pengambilan keputusan dalam bisnis, hingga munculnya inovasi produk dan layanan baru dalam sektor industri (Rick D, 2018). Perkataan ini didukung dengan kesimpulan dari data yang telah dikumpulkan oleh *Deloitte*, bahwa salah satu alasan besar mengapa organisasi kehilangan pertumbuhan dan inovasi adalah karena mereka tidak menggunakan model bisnis yang didukung secara digital (Deloitte, 2020).

Perkembangan teknologi mengubah cara organisasi dalam memiliki kemampuan untuk mengumpulkan, menganalisis, dan memanfaatkan data besar (big data) untuk beroperasi dan membuat keputusan strategis (Alaskar et al., 2024). Penggunaan data telah memungkinkan perusahaan untuk mengidentifikasi pola, membuat prediksi, dan mengoptimalkan berbagai aspek operasional mereka. Seperti yang dikatakan pada jurnal penelitian (Gudivada, 2017) dengan menggunakan data, perusahaan dapat mengambil keputusan yang strategis dan dapat membuat prediksi dari data yang telah dikelola.

Adapun industri yang terdampak adalah industri hiburan, yang meliputi media, film, musik, dan layanan streaming, sangat bergantung pada pemahaman mendalam terhadap perilaku konsumen. Dengan jumlah data yang terus bertambah dari berbagai platform digital, Industri hiburan dapat memanfaatkan data analisis untuk meningkatkan pengalaman pengguna dalam mengembangkan konten yang lebih sesuai dengan preferensi audiens (Feng et al., 2018). Faktanya banyak peneliti telah melakukan penelitian terhadap kasus serupa dan ditemukan bahwa data analisis berdampak signifikan dalam perkembangan pelaku di industri hiburan (Nurhayati, 2019)

Penerapan data analisis ini juga berguna untuk memahami tren pasar agar pemasaran produk lebih efektif dan tepat sasaran. Salah satu metode yang dapat digunakan dalam data analisis pada kasus ini adalah visualisasi data, dimana dengan visualisasi data pihak perusahaan dapat lebih memahami dan mengidentifikasi tentang apa yang terjadi di lapangan (Rachel, 2018).



Gambar 1.1 Jumlah User Berlangganan Di Platform Netflix (Statista, 2022)

Netflix, sebagai salah satu platform terdepan dalam industri hiburan, telah menjadi contoh bagaimana teknologi dan analisis data dapat digunakan untuk mengubah cara orang menonton film dan acara TV. Awalnya Netflix hanya bisnis yang bermodelkan penyewaan DVD hingga akhirnya berkembang sampai menjadi perusahaan besar mancanegara (Sun, 2022) Mengingat diversitas konten yang diproduksi dari berbagai negara, analisis mendalam diperlukan untuk mengeksplorasi bagaimana film dan acara TV dari negara-negara tersebut mempengaruhi preferensi dan keputusan pengguna dalam berlangganan.

Penelitian ini bertujuan untuk mengatasi masalah tersebut dengan menganalisis industri hiburan platform Netflix berdasarkan perbandingan film dan acara TV yang dibuat oleh negara-negara di dunia. Dengan menerapkan analisis data, penelitian ini akan mengeksplorasi bagaimana tayangan film dan acara TV dari berbagai negara mempengaruhi pengguna dalam berlangganan di platform Netflix. Hasil penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam bagi perusahaan dalam mengambil keputusan strategis terkait produksi dan rekomendasi konten yang sesuai dengan preferensi global pengguna.

1.2 Rumusan Masalah

Dengan diversitas konten film dan acara tv yang diproduksi dari berbagai negara diperlukan pemahaman lebih mendalam tentang apa yang membuat pengguna berlangganan di platform. Maka pertanyaan untuk masalah ini adalah “Bagaimana tren dan pola distribusi konten pada platform Netflix dapat diidentifikasi melalui pendekatan eksplorasi data dan klusterisasi menggunakan informasi yang tersedia dalam dataset?”

1.3 Batasan Masalah

Dalam rangka memfokuskan isu yang dibahas, penelitian ini mengadopsi beberapa pembatasan masalah. Beberapa batasan masalah pada penelitian ini adalah:

- a) Data yang diambil berasal dari Kaggle berjumlah sebanyak 8807 data yang berisi 12 atribut perdatanya.
- b) Data diolah menggunakan *Google Colab*.
- c) Data yang digunakan adalah mencakup beberapa variabel, antara lain jenis (film/acara TV), judul, sutradara, aktor, negara, tanggal penambahan, tahun rilis, rating usia, durasi, genre, deskripsi film, *unique id*.
- d) Data yang diambil berisi data dari tahun 2018 - 2021
- e) Metode yang akan digunakan adalah data analisis EDA (*Exploratory Data Analysis*) dan memvisualisasikan data yang telah dikelola untuk diambil kesimpulan.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mengidentifikasi tren dan pola dalam penyajian konten Netflix melalui pendekatan eksploratif berbasis data. Dengan menerapkan teknik eksplorasi data dan klusterisasi, penelitian ini berupaya mengungkap segmentasi serta karakteristik konten yang dapat merepresentasikan dinamika dan perkembangan industri hiburan digital.

1.5 Manfaat Penelitian

- a. Membantu pihak-pihak seperti pengelola bisnis hiburan dalam memahami pola konten, memprediksi perilaku pasar, serta mendukung pengambilan keputusan yang lebih tepat dan strategis.
- b. Membantu sebagai rujukan untuk penelitian dengan topik yang berkaitan.

1.6 Metodologi Penelitian

Metodologi pada penelitian ini akan menggunakan metode data analisis EDA dan visualisasi data, penulis akan menggunakan *Google Colab* sebagai platform dalam membersihkan, menganalisis, serta memvisualisasikan data.



Gambar 1.2 Metode data analisis yang akan digunakan.

- a. *Data Requirement* berisi tentang batasan serta tujuan dalam penelitian ini.
- b. *Data Collection*, langkah yang diperlukan peneliti untuk mengumpulkan data yang relevan, data diambil dari platform kaggle.
- c. *Data Processing*, dilakukan peneliti untuk mengelola data agar ketika di bagian analisis tidak terdapat bias dalam proses.
- d. Analyze data, tahap ini peneliti memutuskan untuk memakai metode apa yang akan digunakan dalam data analisis, disini peneliti menerapkan metode descriptive analisis dan visualisasi data.
- e. *Visualize and interpret result*, akan ditampilkan visualisasi terhadap data yang tadinya berbentuk angka maupun kata menjadi berbentuk diagram, plot, dan chart.

BAB II

KAJIAN PUSTAKA

2.1 Penelitian Terdahulu

Pada bagian ini, peneliti akan menyajikan hasil dari beberapa penelitian yang telah dilakukan sebelumnya dengan topik yang relevan. Pemaparan hasil penelitian sebelumnya menjadi aspek yang penting karena dapat digunakan sebagai referensi dalam penelitian ini.

Penelitian	Judul Penelitian	Industri	Metode	Hasil Penelitian
(Iskandar & Sutanto, 2022)	<i>Analysis of Movie Genre Availability, Service Quality, and Movie Quality on Netflix User Satisfaction in Indonesia</i>	Industri Hiburan	<i>multiple linear regression with the SPSS program</i>	Data analisis mendapatkan temuan penelitian ini menunjukkan bahwa ketersediaan genre film dinyatakan berpengaruh secara tidak signifikan terhadap kepuasan pelanggan, adapun kualitas layanan Netflix dinyatakan berpengaruh secara signifikan terhadap kepuasan, dan kualitas film Netflix dinyatakan berpengaruh secara signifikan terhadap kepuasan pelanggan.
(Anggrahini et al., 2020)	Implementasi Data Analitik Untuk Penentuan Jadwal Pemeliharaan Mesin Pada Perusahaan Manufaktur Kontinu Skala Besar	Manufaktur	K-means <i>Clustering</i> dengan perhitungan <i>Elbow method</i>	Data analitik membantu dalam menentukan jadwal pemeliharaan mesin yang lebih efisien, baik secara terus menerus atau berdasarkan kebutuhan terhadap mesin.

(Hermawan, 2022)	<i>Analysis of interest of the YouTube platform users as a medium of digital communication today</i>	Industri Hiburan	Survei dan analisis data sekunder	Mendapatkan hasil berupa minat para pengguna YouTube terkait dengan muatan video maupun teknik operasi yang tersedia di platform ini.
(Anisa & Novita, 2023)	Penggunaan Data Analytics Terhadap Peningkatan Deteksi Kecurangan Pada Pemeriksaan Eksternal	Jasa keuangan	<i>Data analisis Structural Equation Modeling (SEM)</i>	Penggunaan data analytics berpengaruh positif dan signifikan terhadap pendeteksian kecurangan. Data analytics membantu auditor dalam mendeteksi indikasi kecurangan dengan cepat.
(Fernandes Andry et al., 2021)	<i>BIG DATA ANALYSIS ON YOUTUBE WITH TABLEAU</i>	Industri Hiburan	<i>Data Visualization Tableau</i>	Visualisasi data analisis dapat membantu YouTube dan penggunanya mengidentifikasi area mana yang perlu ditingkatkan, faktor-faktor apa yang mempengaruhi kepuasan dan ketidakpuasan pengguna, serta apa yang harus dilakukan dan diketahui untuk mengembangkan bagian yang kurang dan memperhatikan kebutuhan pengguna saat ini.
(Mulyandi et al., 2021)	<i>Analysis of Disney plus Hotstar Indonesia Youtube Advertising Version "Choose All, For All" During the Covid 19</i>	Industri Hiburan	Data analisis statistical analisis	Dari hasil penelitian ini, dapat dilihat bahwa ada satu indikator yang tidak berpengaruh, yaitu indikator konten yang mengganggu konsumen. Oleh karena itu, Disney Hotstar perlu lebih memperhatikan

	<i>Pandemic in July to December 2020</i>			penyajian iklan video di YouTube agar dapat memasang fitur konten iklan yang dapat dilewati oleh penonton.
(Celicia et al., 2022)	Penerapan Data Science pada Analisis Data Acara TV dan Film pada Aplikasi Layanan Streaming	Industri Hiburan	<i>Clustering & N-gram, EDA (Exploratory Data Analysis),</i> sistem rekomendasi, <i>Cosine Similarity, K-means</i>	Analisis data dilakukan untuk menampilkan rekomendasi untuk sebuah tayangan di Netflix menggunakan algoritma <i>Cosine Similarity</i> , menerapkan <i>Clustering</i> menggunakan K-means, dan <i>N-Gram</i> dari hasil <i>Clustering</i> .
(Anggraini et al., 2023)	Analisis PIECES untuk Evaluasi Layanan Aplikasi Disney+ Hotstar	Industri Hiburan	Data analisis <i>descriptive analysis</i>	Dengan Data analisis mendapatkan hasil berupa keunggulan aplikasi Disney+ Hotstar yaitu layanan yang diberikan pada aspek kinerja aplikasi, informasi yang disajikan, efisiensi dan layanan pelanggan. sedangkan layanan pada aspek <i>Economics</i> dan <i>Control and Security</i> masih perlu ditingkatkan agar menjadi sangat puas pada keseluruhan aspek PIECES

Dari tabel penelitian di atas, dapat disimpulkan bahwa data analisis telah diterapkan secara luas di berbagai industri untuk meningkatkan kinerja dan efisiensi. Contohnya dalam industri jasa keuangan, penelitian oleh (Anisa & Novita, 2023) menunjukkan bahwa penggunaan data analisis melalui metode *Structural Equation Modeling (SEM)* secara signifikan meningkatkan kemampuan auditor dalam mendeteksi indikasi kecurangan dengan cepat dan akurat.

Di industri hiburan sendiri, sejumlah penelitian menunjukkan berbagai aplikasi data analisis untuk meningkatkan layanan dan memahami perilaku pengguna. salah satunya (Hermawan, 2022) melalui survei dan analisis data sekunder mendapatkan temuan minat pengguna YouTube terhadap muatan video maupun teknik operasi yang tersedia di platform ini.

Adapun penelitian lainnya tentang data analisis pada industri hiburan pada aplikasi layanan streaming Netflix, menunjukkan bahwa algoritma seperti *Cosine Similarity* dan K-means dapat digunakan untuk mendapatkan analisis tentang cara meningkatkan rekomendasi tayangan di platform.

Secara keseluruhan, penelitian-penelitian ini menunjukkan bahwa data analisis mampu memberikan kontribusi signifikan dalam meningkatkan efisiensi operasional, memahami perilaku pengguna, dan meningkatkan kualitas layanan di berbagai industri. Referensi ini akan peneliti gunakan untuk menerapkan data analisis di industri hiburan yaitu Netflix dengan harapan mendapatkan output berupa kesimpulan dari data yang akan dikelola.

Penelitian ini lebih mengutamakan analisis terkait tren suatu film dan movie yang dirilis oleh Netflix, Metode yang akan digunakan dalam penelitian ini berupa EDA (*Exploratory Data Analyst*) dan K-means untuk mengidentifikasi tren suatu film dan acara tv.

2.2 Data Analisis

Data analisis adalah proses pengorganisasian dalam memilah data ke dalam pola, kategori, dan satuan uraian dasar (Nasir & Sukmawati, 2023). Dalam era di mana data dianalogikan sebagai minyak, Data analisis menjadi landasan untuk mengintegrasikan data yang heterogen dari berbagai sumber, lalu menarik kesimpulan darinya dan membuat prediksi untuk keperluan penyusunan inovasi, peningkatan keunggulan bisnis dan pembuatan keputusan strategis (Anggrahini et al., 2020; Gudivada, 2017)). Maka dari itu pentingnya data analisis tidak bisa diabaikan dalam berbagai sektor kehidupan. Di sektor industri hiburan seperti Netflix, analisis data membantu perusahaan untuk memahami perilaku konsumen, mengidentifikasi tren pasar, dan mengoptimalkan operasional. Dengan data yang dianalisis dengan baik, perusahaan dapat membuat keputusan yang lebih tepat dan meningkatkan efisiensi serta profitabilitas mereka.

2.3 Exploratory Data Analysis (EDA)

Pendekatan dalam analisis data yang bertujuan untuk memahami struktur, pola, dan karakteristik utama dari suatu dataset dan memodelkannya dalam bentuk visualisasi data (Sahoo* et al., 2019). Konsep ini diperkenalkan oleh John W. Tukey pada tahun 1977 sebagai respons terhadap ketergantungan yang berlebihan pada pengujian hipotesis formal dalam statistik tradisional. EDA menekankan pentingnya eksplorasi data secara visual dan deskriptif untuk mengidentifikasi anomali, outlier, serta hubungan antar variabel yang mungkin tidak terdeteksi melalui pendekatan statistik konvensional.

2.4 K-means

Algoritma *unsupervised learning* yang digunakan untuk mengelompokkan data ke dalam sejumlah kluster berdasarkan kesamaan fitur atau atribut. Clustering digunakan dalam eksplorasi dataset untuk melihat pola ataupun karakteristik (Dangeti, 2017). Algoritma ini bekerja dengan membagi data ke dalam k kluster, di mana setiap data akan dimasukkan ke dalam kluster dengan centroid terdekat (Mulyati et al., 2025). Proses ini dilakukan secara iteratif hingga posisi centroid stabil atau mencapai jumlah iterasi maksimum berikut rumus perhitungannya:

$$\sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 \quad (1)$$

2.5 Elbow Method

Pendekatan heuristik yang digunakan untuk menentukan jumlah kluster optimal (k) dalam algoritma *K-Means Clustering*. Metode ini melibatkan analisis grafis terhadap nilai *Within-Cluster Sum of Squares* (WCSS) untuk berbagai nilai k (Jollyta et al., 2019). Dengan memplot nilai WCSS terhadap jumlah kluster, akan terbentuk grafik yang menyerupai siku (*elbow*). Titik di mana penurunan WCSS mulai melambat secara signifikan dianggap sebagai jumlah kluster yang optimal, karena penambahan kluster setelah titik tersebut tidak memberikan pengurangan WCSS yang substansial. Pengujian K-Means Clustering dapat dilakukan dengan menghitung Sum of Squared Errors (SSE) (Mulyati et al., 2025) rumus perhitungannya sebagai berikut:

$$SSE = \sum_{i=1}^k \sum_{x \in S_i} (x - \mu_i)^2 \quad (2)$$

2.6 Industri Hiburan

Mendapatkan penonton baru dan mempertahankan penonton lama dengan menjaga mereka tetap terlibat dalam platform, menjadi suatu tantangan terbesar bagi pelaku industri hiburan di seluruh dunia. Dengan mengintegrasikan analisis data, perusahaan industri hiburan mampu memperoleh wawasan mendetail mengenai penonton mereka dari produk yang mereka sukai hingga proses mereka dalam platform hingga menemui tontonan yang disukai (SGA Knowledge Team, 2023). Dengan pemahaman mendalam terhadap data, perusahaan mampu merancang strategi lebih efisien dan responsif terhadap perubahan pasar.

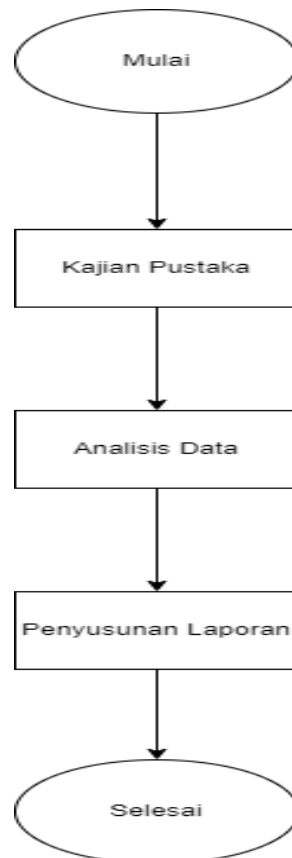
BAB III

METODE PENELITIAN

Berikut adalah tahapan metodologi penelitian yang digunakan dalam proses ini:

1. **Start** – Pada tahap awal, peneliti menetapkan topik dan tujuan penelitian serta merumuskan masalah yang akan dibahas. Pemilihan metode penelitian juga dilakukan pada tahap ini untuk memastikan langkah-langkah penelitian berjalan secara sistematis.
2. **Kajian Pustaka** – Selanjutnya, peneliti mengumpulkan dan menelaah literatur atau referensi dari berbagai sumber terpercaya. Kajian pustaka membantu memberikan dasar teori dan mendukung argumen yang relevan dengan penelitian.
3. **Analisis Data** – Setelah data terkumpul, peneliti melakukan proses analisis menggunakan *Google Colab*, lalu melakukan visualisasi data agar tren yang ingin disampaikan dapat dipahami oleh orang awam. Analisis EDA dan K-means ini bertujuan untuk mengidentifikasi pola atau hubungan yang signifikan dalam data.
4. **Menyusun Laporan** – Tahap ini melibatkan penyusunan hasil penelitian secara sistematis dalam bentuk laporan. Peneliti menyajikan temuan, interpretasi, dan kesimpulan yang diperoleh dari analisis data, disertai rekomendasi bila diperlukan.
5. **Selesai** – Penelitian diakhiri dengan tahap finalisasi, yang mencakup validasi hasil dan presentasi laporan kepada pihak terkait, seperti dosen pembimbing atau pihak universitas.

Setiap tahapan ini berperan penting dalam menjaga kualitas penelitian, dengan analisis data sebagai inti untuk menarik kesimpulan yang akurat dan relevan.



Gambar 3.1 metodologi penelitian yang digunakan dalam proses ini

Data merupakan suatu komponen penting di perkembangan zaman sekarang, Hal ini pernah diungkapkan oleh matematikawan dari Inggris yang bernama Clive Humby, dia pernah berkata “*Data is new oil*”, yang menandakan bahwa setiap aktivitas hidup kita telah saling terkait dengan data. Pada bab ini akan dijabarkan penerapan analisis data pada Platform Netflix menggunakan python di *Google Colab*, dalam menerapkan data analisis terdapat beberapa tahap yang akan dilakukan (lihat gambar 3.1) sebagai berikut: 1) *Data Requirement*. 2) *Data Collection*. 3) *Data Processing*. 4) *Data Analysis*. 5) *Visualize and interpret the result* (Arora, 2020).



Gambar 3.2 Langkah metode analisis data

3.1 Data Requirement

Penting dasarnya dalam melakukan pendefinisian pertanyaan sebagai acuan dalam tujuan agar tidak keluar dari topik yang telah ditetapkan, pertanyaan yang ingin dibahas dalam tahap ini adalah: "Bagaimana tren dan pola distribusi konten pada platform Netflix dapat diidentifikasi melalui pendekatan eksplorasi data dan klusterisasi menggunakan informasi yang tersedia dalam dataset?".

Dalam konteks ini, kita ingin memahami bagaimana distribusi film dan acara tv bervariasi di antara negara, apakah ada perbedaan tren atau preferensi dalam pembuatan dan konsumsi konten dari satu negara ke negara lainnya, dan apakah terdapat pola-pola geografis yang dapat diidentifikasi. Untuk menjawab pertanyaan mari kita lakukan pembahasannya.

3.2 Data Collection

Data yang akan digunakan dalam penelitian ini akan diambil dari sumber eksternal, yaitu platform Kaggle. Data Netflix yang akan digunakan akan mencakup beberapa variabel, antara lain jenis (film/acara TV), judul, sutradara, pemeran, negara, tanggal penambahan, tahun rilis, dan jumlah pengguna per tahun. Data ini akan disediakan dalam format CSV untuk memudahkan analisis. Data berjumlah 8807 yang berisi 12 kolom perdatanya, data memiliki rentang dari tahun 2018 - 2021.

3.3 Data Processing

Pembersihan data merupakan salah satu langkah yang paling penting dalam analisis data karena jika data yang digunakan tidak konsisten dan akurat maka akan berakibat fatal terhadap hasil analisis. Fokus utama dari pembersihan data adalah untuk memverifikasi bahwa data yang dipakai dalam analisis adalah tepat, sesuai format, dan terbebas dari kesalahan.

Tugas utama pembersihan data meliputi:

1. Menghapus data duplikat dan outlier.
2. Menghapus titik data yang tidak diinginkan, mengabaikan data yang tidak relevan dengan tujuan analisis.
3. Memastikan dan mengubah format data yang sesuai agar mudah diolah pada tahap data analisis dan visualisasi.
4. Mengisi dan menghapus data yang tidak memiliki nilai.

Pada penelitian ini data yang tidak diinginkan dan tidak relevan merujuk pada data yang secara logis tidak masuk akal atau bertentangan dengan kenyataan, sehingga dapat merusak hasil analisis. Contohnya dalam penelitian ini adalah data dengan nilai time negatif (yaitu tahun ditambahkan konten ke Netflix lebih awal dari tahun rilis konten) misal tahun film ditambahkan ke netflix pada 2020 akan tetapi rilisnya film tersebut pada 2022. Hal ini tidak mungkin terjadi secara logika, sehingga dianggap sebagai kesalahan input data dan perlu dihapus.

Awalnya data memiliki format tanggal di kolom `date_added` menggunakan format teks panjang, contohnya: "September 24, 2021". Format ini tidak sesuai untuk pengolahan waktu secara langsung dalam Python, khususnya saat menggunakan *library* seperti `pandas` yang membutuhkan format tanggal baku agar bisa diproses lebih lanjut. Oleh karena itu, dilakukan konversi format tanggal tersebut ke dalam format standar ISO 8601, yaitu "YYYY-MM-DD" (contoh: "2021-09-24"). Format ini lebih kompatibel dengan proses analisis data dan visualisasi berbasis waktu.

3.4 Data Analysis

Data Analysis dilakukan dengan metode EDA (*Exploratory Data Analysis*), EDA adalah proses pemahaman dan pembersihan data untuk memastikan kualitas data siap dianalisis lebih lanjut. EDA meliputi eksplorasi struktur data, deteksi missing values dan outlier, analisis distribusi variabel secara univariat dan bivariat, serta pembuatan visualisasi data untuk mengidentifikasi pola dan hubungan antar fitur. Proses ini sebagai pondasi awal sebelum dilakukan pemodelan atau pengambilan keputusan berdasarkan data.

Adapun penjelasan tahapan yang dilakukan :

a. Memahami Struktur Data

Langkah pertama dalam *Exploratory Data Analysis (EDA)* adalah memahami struktur data yang digunakan. Proses ini mencakup pengamatan terhadap dimensi dataset, seperti jumlah baris dan kolom, Tahap ini penting untuk mengidentifikasi tipe data dari masing-masing kolom, seperti string, integer, float, atau datetime, guna memastikan setiap atribut diproses sesuai dengan karakteristiknya.

b. Analisis Univariate

Analisis univariate dilakukan untuk memahami karakteristik distribusi dari setiap variabel secara individu. Untuk data bertipe numerik, analisis ini dilakukan dengan deskripsi statistik seperti nilai rata-rata, minimum dan maksimum. Sementara itu, untuk data kategorikal, digunakan visualisasi berupa bar chart untuk mengetahui dominasi masing-masing kategori.

c. Analisis Bivariate

Analisis bivariate bertujuan untuk mengevaluasi hubungan antara dua variabel. Adapun hasilnya berupa temuan seperti korelasi jenis tayangan film yang banyak dirilis pada bulan Juli (565 rilis), sedangkan acara TV lebih sering dirilis pada bulan Desember (265 rilis).

3.5 Visualize and interpret the result**A. Visualisasi Data**

Di tahap terakhir ini akan ditampilkan visualisasi hasil temuan analisis pada metode EDA, Visualisasi data merupakan teknik yang digunakan untuk mewakili informasi dalam bentuk grafik atau diagram, sehingga memudahkan pemahaman pola-pola dan hubungan antar variabel. Pentingnya visualisasi dapat membantu pembaca melihat dan memahami data dengan lebih baik. Baik itu sederhana maupun kompleks, visualisasi yang tepat mampu membawa semua pembacanya ke arah yang sama tanpa melihat tingkat keahlian mereka.

B. K-means Clustering

Dalam penelitian ini digunakan algoritma K-means *Clustering* untuk mengelompokkan data berdasarkan kemiripan karakteristik. Proses *Clustering* dimulai dengan menentukan jumlah cluster (K), diikuti dengan inisialisasi centroid secara acak. Selanjutnya, setiap data dihitung jaraknya ke centroid dan dikelompokkan ke dalam cluster terdekat. Centroid akan diperbarui berdasarkan rata-rata data di dalam cluster dan proses ini diulang hingga konvergen. Penentuan jumlah K optimal dilakukan menggunakan *Elbow method* dan *Silhouette score*.

a) *Elbow method*

Elbow method digunakan untuk menentukan nilai K optimal dengan cara memplot nilai *Within-Cluster Sum of Squares* (WCSS) terhadap berbagai nilai K. Grafik akan membentuk kurva menurun, dan titik "siku" (*elbow*) menunjukkan K yang optimal yaitu saat penurunan WCSS mulai melambat.

b) *Silhouette score*

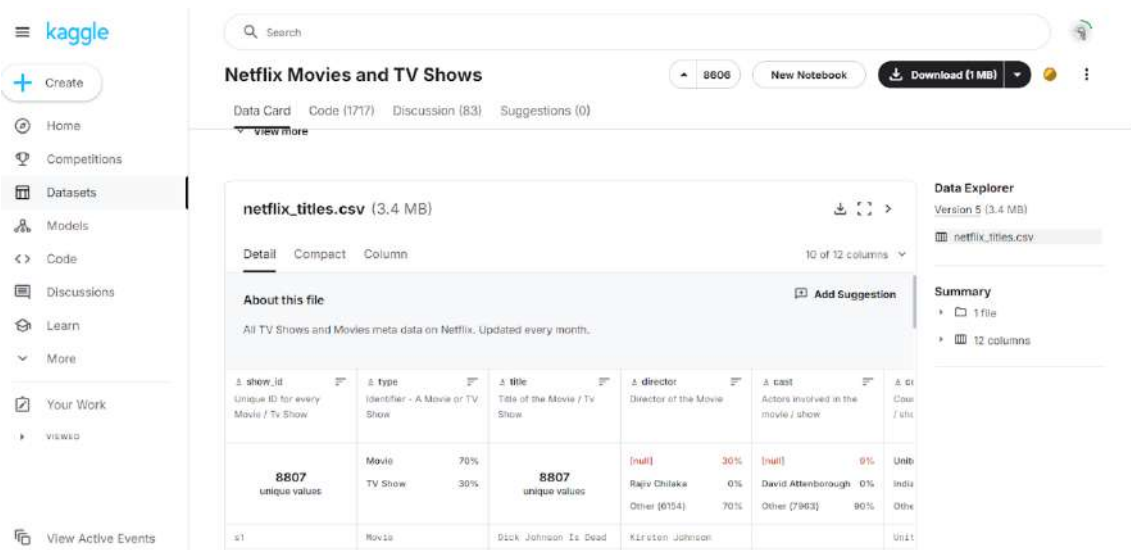
Silhouette score mengukur seberapa mirip suatu data dengan cluster-nya sendiri dibandingkan dengan cluster lain. Nilai *Silhouette score* berkisar antara -1 sampai 1. Nilai mendekati 1 menunjukkan data berada di cluster yang benar. Nilai mendekati 0 berarti data berada di perbatasan antara dua cluster. Nilai negatif berarti data bisa saja lebih cocok berada di cluster lain.

Penggunaan K-means *Clustering* membantu mengidentifikasi kelompok film dan acara tv dengan pola produksi konten yang serupa, memberikan wawasan tentang strategi produksi konten di berbagai wilayah geografis.

BAB IV HASIL DAN PEMBAHASAN

4.1 Data Collection

Data yang digunakan pada penelitian diperoleh melalui kaggle, Informasi terkait data bisa diakses dengan mengklik bagian dataset pada kolom pencarian, Seperti yang ditunjukkan pada Gambar 4.1



Gambar 4.1 Halaman Netflix Dataset

Pada Gambar 4.1 telah terlihat beberapa rincian awal dari data, dimulai dari format data yaitu csv, jumlah dataset Netflix yang memuat 8807 baris data yang berisikan jenis film atau acara tv beserta rinciannya yang berjumlah 12 kolom berbeda. Detail data terlampir pada Gambar 4.2.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description				
s1	Movie	Dick Johnson	Kristen Johnson	United States	2020	PG-13	90 min	Documentary	As her father nears the end of his life, filmmaker Kristin Johnson						
s2	TV Show	Blood & Water	Anna Qamisa	South Africa	2021	TV-MA	2 Seasons	International	After crossing paths at a party, a Cape Town teen sets out to						
s3	TV Show	Ganglands	Julien Leck	Sami Bouajila, Tracy Spiridakis	2021	TV-MA	3 Seasons	Crime TV	To protect his family from a powerful drug lord, a young man						
s4	TV Show	Jaibirds	New Orleans		2021	TV-MA	3 Seasons	Documentary	Feuds, flirtations and toilet talk go down among the						
s5	TV Show	Kota Factory	Mayur Mohan	India	2021	TV-MA	2 Seasons	International	In a city of coaching centers known to train India's elite, a						
s6	TV Show	Midnight & Mike	Flannery Kate Siegel, Zach Gilford		2021	TV-MA	3 Seasons	TV Drama	The arrival of a charismatic young priest brings glory to						
s7	Movie	My Little P	Robert Cul	Vanessa Hudgens, Kin Shriner	2021	PG	91 min	Children & Equestria	is divided. But a bright-eyed hero believes in the						
s8	Movie	Sankofa	Halle Geri	Kofi Ghannem	United States	1993	TV-MA	125 min	Drama	In On a photo shoot in Ghana, an American model slips					
s9	TV Show	The Great	Andy Deitch	Mel Giedroyc	United Kingdom	2021	TV-14	9 Seasons	British TV	A talented batch of amateur bakers face off in a 10-w					
s10	Movie	The Starlin	Theodore Melissa M	United States	2021	PG-13	104 min	Comedies	A woman adjusting to life after a loss contends with a						
s11	TV Show	Vendetta: Truth, Lies and The Mafia			2021	TV-MA	3 Seasons	Crime TV	Sicily boasts a bold "Anti-Mafia" coalition. But what h						
s12	TV Show	Bangkok 8	Kongkiat K	Sukollawat Kanarot, Sittichai	2021	TV-MA	3 Seasons	Crime TV	Struggling to earn a living in Bangkok, a man joins an e						
s13	Movie	Je Suis Kar	Christian S	Luna Wedel	Germany, India	2021	TV-MA	127 min	Drama	In After most of her family is murdered in a terrorist bon					
s14	Movie	Confessor	Bruno Gar	Klara Castanho, Lucca	2021	TV-PG	91 min	Children & When	the clever but socially awkward Teru joins a n						
s15	TV Show	Crime Stories: India	Detectives		2021	TV-MA	3 Seasons	British TV	Cameras following Bengaluru police on the job offer a						
s16	TV Show	Dear White	People Logan Bror	United States	2021	TV-MA	4 Seasons	TV Comedy	Students of color navigate the daily slights and slipper						
s17	Movie	Europe's & Pedro	de Echave	García-a, Pablo	2020	TV-MA	67 min	Documentary	Declassified documents reveal the post-WWII life of C						
s18	TV Show	Falsa	Identidad Luis Ernest	Mexico	2020	TV-MA	2 Seasons	Crime TV	Strangers Diego and Isabel flee their home in Mexico i						
s19	Movie	Intrusion	Adam Salk	Freida Pinto, Logan M	2021	TV-14	94 min	Thrillers	After a deadly home invasion at a couple's new dr						
s20	TV Show	Jaguar	Blanca Suárez	Spain	2021	TV-MA	3 Seasons	International	In the 1960s, a Holocaust survivor joins a group of self						
s21	TV Show	Monsters	Olivier Megaton		2021	TV-14	3 Seasons	Crime TV	In the late 1970s, an accused serial rapist claims multi						
s22	TV Show	Resurrection: Erbu	Engin Altan	Turkey	2018	TV-14	5 Seasons	International	When a good deed unwittingly endangers his clan, a I						
s23	Movie	Avvai Shan	K.S. Raviki, Kamal	Hassan, Meenu	1996	TV-PG	161 min	Comedies	Newly divorced and denied visitation rights with his d						
s24	Movie	Gol Gol	Cr Alex Woo, Maisie	Benson, Pascal	2021	TV-Y	61 min	Children & From	arcade games to sled days and hiccup cures, Con						
s25	Movie	Jehans	S. Shankar	Prashanthi	India	1998	TV-14	166 min	Comedies	When the father of the man she loves insists that his t					
s26	TV Show	Love on the	Spectrum	Brooke Sar	Australia	2021	TV-14	2 Seasons	Documentary	Find love can be hard for anyone. For young adults					

Gambar 4.2 Netflix_titles.csv

Setiap kolom data memiliki rincian tersendiri yang akan dijelaskan pada Tabel 4.1

Tabel 4.1 Tabel rincian kolom data

Kolom	Deskripsi
<i>show_id</i>	<i>Unique ID for every Movie / Tv Show</i>
<i>title</i>	<i>Title of the Movie / Tv Show</i>
<i>director</i>	<i>Director of the Movie</i>
<i>cast</i>	<i>Actors involved in the movie / show</i>
<i>country</i>	<i>Country where the movie / show was produced</i>
<i>date_added</i>	<i>Date it was added on Netflix</i>
<i>release_year</i>	<i>Actual Release year of the move / show</i>
<i>rating</i>	<i>TV Rating of the movie / show</i>
<i>duration</i>	<i>Total Duration - in minutes or number of seasons</i>
<i>listed_in</i>	<i>Genre</i>
<i>description</i>	<i>The summary description</i>
<i>type</i>	<i>Identifier - A Movie or TV Show</i>

Dataset yang digunakan dalam penelitian ini diambil dari platform *Kaggle* dan berjudul "netflix_titles", yang mencakup sebanyak 8807 baris data dengan 12 atribut utama. Data ini memuat informasi detail mengenai film dan acara TV yang tersedia di platform Netflix hingga tahun 2021. Beberapa atribut penting yang tersedia meliputi: *show_id*, *title* (judul tayangan), *type* (jenis tayangan: Movie atau TV Show), *country* (negara produksi), *release_year* (tahun rilis), *duration* (durasi film atau jumlah season), *director*, *cast*, *listed_in* (genre), *description* (deskripsi) *rating* (usia), hingga *date_added* (tanggal ditambahkan ke platform).<

Dataset ini cocok untuk digunakan dalam analisis segmentasi konten karena menyediakan kombinasi variabel kategorikal dan numerik yang dapat dieksplorasi lebih lanjut, seperti perbandingan tren produksi berdasarkan waktu hingga genre. Informasi ini penting dalam konteks industri hiburan karena dapat membantu memahami strategi produksi dan distribusi konten oleh Netflix secara lebih mendalam.

4.2 Data Processing

Dalam tahap *data cleaning*, fokusnya adalah memastikan kualitas data agar lebih akurat dan siap digunakan dalam analisis. Data yang bersih membantu menghindari bias atau kesalahan dalam hasil analisis. Setelah mengambil data dari kaggle teridentifikasi adanya kejanggalan dalam data seperti adanya nilai yang kosong didata, contoh dari data yang mengalami kejanggalan bisa dilihat di Gambar 4.3.



show_id	0
type	0
title	0
director	2634
cast	825
country	831
date_added	10
release_year	0
rating	4
duration	3
listed_in	0
description	0

Gambar 4.3 Data Tidak Memiliki Nilai

Data yang diambil masih memiliki format tanggal yang masih belum sesuai dengan operasi yang akan dijalankan, Format penanggalan data yang kurang tepat bisa dilihat di Gambar 4.4.

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
s2	TV Show	Blood & Water	NaN	Ama Camata, Khosi Ngema, Gail Mablane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town...
s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train L...
s11	TV Show	Vendetta: Truth, Lies and The Mafia	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, Docuseries, International TV S...	Sicily boasts a bold "Anti-Mafia" coalition. B...
s15	TV Show	Crime Stories: India Detectives	NaN	NaN	NaN	September 22, 2021	2021	TV-MA	1 Season	British TV Shows, Crime TV Shows, Docuseries	Cameras following Bengaluru police on the job ...

Gambar 4.4 Format Penanggalan Belum Sesuai

Pada data juga dilakukan pengecekan apakah data memiliki data duplikat atau tidak, data duplikat harus diidentifikasi dan dihapus dikarenakan dapat mengganggu hasil dari analisis data.

```
print("Jumlah duplikasi: ", net_df.duplicated().sum())
```

Gambar 4.5 Kode Pengecekan Duplikasi Data

Penyesuaian data menggunakan python di *Google Colab*, pada gambar 4.5 akan dilakukan penggantian dan menghapus nilai null yang telah diidentifikasi dan pengolahan data tanggal menggunakan fungsi `pd.to_datetime()`. Perubahan ini penting dikarenakan kita akan melakukan operasi pada data tanggal.

```
net_df["date_added"] = net_df["date_added"].str.strip()
net_df["date_added"] = pd.to_datetime(net_df['date_added'])
net_df['month_added'] = net_df['date_added'].dt.month
net_df['month_name_added'] = net_df['date_added'].dt.month_name()
net_df['year_added'] = net_df['date_added'].dt.year

net_df['country'] = net_df['country'].fillna(net_df['country'].mode()[0])
net_df['cast'].replace(np.nan, 'No Data', inplace = True)
net_df['director'].replace(np.nan, 'No Data', inplace = True)
net_df.dropna(inplace=True)
net_df.drop_duplicates(inplace= True)
```

Gambar 4.6 Kode Pembersihan Data

Pada dataset ditemukan beberapa kolom yang memiliki nilai kosong, seperti *country*, *cast*, dan *director*. Untuk kolom *country*, dilakukan pengisian nilai kosong menggunakan modus (nilai yang paling sering muncul) dengan fungsi *fillna()* dan *mode()*, karena diasumsikan bahwa negara yang paling sering muncul mewakili mayoritas data untuk dianalisis. Sementara itu, untuk kolom *cast* dan *director*, nilai kosong tidak diisi dengan data asli melainkan diganti dengan string "No Data" menggunakan fungsi *replace()*. Hal ini dilakukan untuk mempertahankan data tersebut dalam analisis tanpa menghilangkannya, namun tetap memberi informasi bahwa data tersebut tidak tersedia.

Setelah mengisi nilai kosong pada kolom-kolom tertentu, baris yang masih memiliki nilai kosong di kolom lainnya dianggap tidak dapat digunakan dalam analisis lanjutan dan dihapus dengan fungsi *dropna()*. Langkah ini dilakukan agar data yang digunakan dalam analisis bersih dan tidak menimbulkan bias atau error dalam visualisasi dan perhitungan. Selain itu, dilakukan juga penghapusan data duplikat menggunakan fungsi *drop_duplicates()* untuk memastikan bahwa tidak ada entri yang berulang.

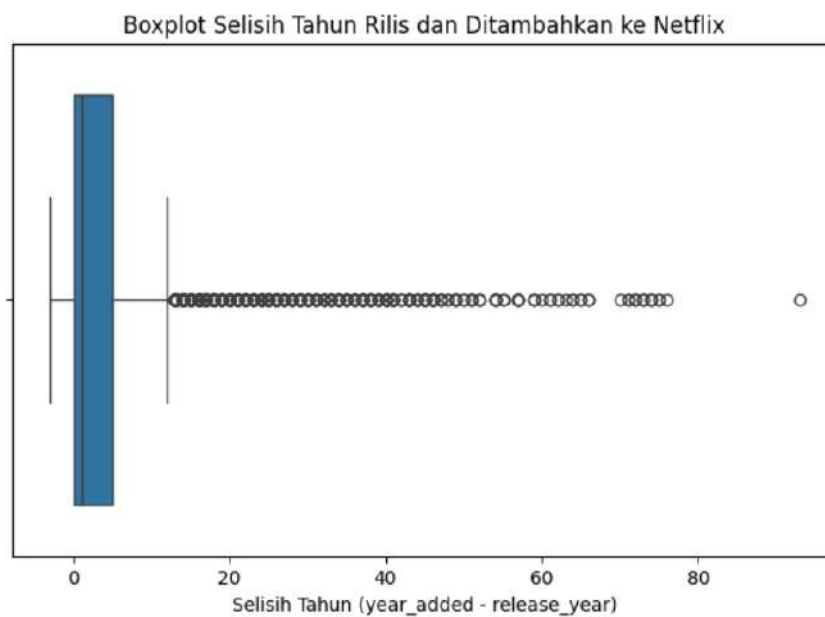
Awalnya format tanggal pada kolom data *date_added* menggunakan format teks, contohnya: "September 24, 2021". Format ini tidak dapat langsung digunakan untuk analisis waktu dalam Python, karena masih dalam bentuk string, khususnya saat menggunakan *library* seperti *pandas* yang membutuhkan format tanggal baku agar bisa diproses lebih lanjut. Oleh karena itu, dilakukan konversi format tanggal tersebut ke dalam format standar ISO 8601, yaitu "YYYY-MM-DD" (contoh: "2021-09-24"). langkah pertama yang dilakukan adalah membersihkan data tersebut dari spasi yang tidak diperlukan menggunakan fungsi *str.strip()*. Selanjutnya, data dikonversi ke dalam format *datetime* menggunakan fungsi *pd.to_datetime()*. Dengan perubahan ini, data tanggal dapat dikenali oleh sistem sebagai tipe data waktu yang valid. Format ini lebih kompatibel dengan proses analisis data dan visualisasi berbasis waktu.

Setelah dilakukannya penyesuaian data, dapat dilihat bahwa format seperti penanggalan dan data yang tidak memiliki nilai telah diubah menjadi format yang dapat digunakan untuk ketahap selanjutnya. Perbedaan data yang telah diubah dapat dilihat di gambar 4.6.

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	month_added	month
0	s1	TV Show	3%	No Data	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	2020-08-14	2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit an island ...	8
1	s2	Movie	7:19	Jorge Michel Grau	Demian Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	2016-12-23	2016	TV-MA	93 min	Dramas, International Movies	After a devastating earthquake hits Mexico Cit...	12

Gambar 4.7 Hasil penyesuaian format data

Dalam mengidentifikasi data outlier dan data yang tidak relevan, langkah awal yang dilakukan adalah dengan memvisualisasikannya menggunakan boxplot. Boxplot merupakan salah satu jenis visualisasi statistik yang efektif untuk menampilkan distribusi data serta mengidentifikasi titik-titik data yang berada di luar batas normal, yang ditampilkan sebagai titik-titik terpisah dari kotak utama.

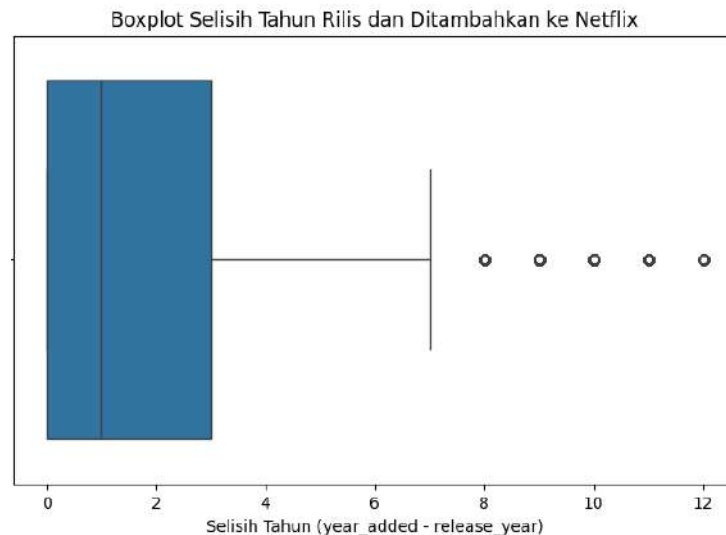


Gambar 4.8 Boxplot

Ditemukan data dengan nilai *time* negatif, *time* adalah variabel dari (tahun ditambakkannya konten ke Netflix – tahun rilisnya konten). *time* akan digunakan untuk analisis rata-rata penambahan konten ke Netflix pada subbab 4.3 Data Analysis.

Secara logis, sebuah konten tidak mungkin ditambahkan ke platform Netflix sebelum tanggal perilisannya. Misalnya, terdapat kasus di mana sebuah film tercatat ditambahkan ke

Netflix pada tahun 2020, namun memiliki tahun rilis 2023. Kondisi seperti ini menunjukkan adanya inkonsistensi data. Oleh karena itu, data dengan nilai time kurang dari 0 dianggap sebagai kesalahan atau anomali, karena tidak sesuai dengan urutan waktu yang seharusnya. Data tersebut dihapus dari dataset guna menjaga kualitas dan validitas hasil analisis yang akan dilakukan..



Gambar 4.9 Boxplot yang telah dilakukan IQR

Setelah proses pembersihan dilakukan, yaitu dengan menghapus data yang memiliki selisih tahun negatif serta data yang berada di luar batas IQR, diperoleh hasil data yang lebih bersih dan relevan seperti terlihat pada Gambar 2. Pada gambar tersebut, tidak ditemukan lagi nilai selisih tahun yang negatif dan persebaran data menjadi lebih terkonsentrasi di sekitar nilai tengah (median), yang menunjukkan bahwa data telah berada dalam rentang yang wajar dan dapat diandalkan untuk analisis selanjutnya.

Tahapan yang dilakukan untuk mengidentifikasi dan menghapus nilai-nilai ekstrem (*outlier*) yang dapat mempengaruhi hasil analisis, digunakan metode Interquartile Range (IQR). Metode IQR bekerja dengan menghitung rentang antara kuartil pertama (Q1) dan kuartil ketiga (Q3), kemudian menetapkan batas bawah dan atas (yaitu $Q1 - 1.5 \times IQR$ dan $Q3 + 1.5 \times IQR$). Data yang berada di luar batas tersebut dianggap sebagai *outlier* dan dihapus dari dataset. Proses ini penting untuk memastikan distribusi data lebih bersih dan representatif terhadap kondisi mayoritas.

4.3 Data Analysis

Data Analysis merupakan salah satu kegiatan dalam menganalisis data yang masih mentah dan diproses dengan cara-cara tertentu sehingga menjadi kesimpulan yang bisa dijadikan referensi untuk pengambilan keputusan, tentunya kita akan melakukan EDA dalam penelitian ini. Dalam EDA, kita melakukan sejumlah aktivitas, seperti membuat visualisasi, mengidentifikasi trend dalam data, dan mengecek apakah ada data yang hilang atau tidak konsisten. Pengecekan data telah dilakukan diawal dan data juga telah dibersihkan. Maka langkah selanjutnya adalah eksplorasi data (EDA).

```
unique_films = net_df[net_df['type'] == 'Film']['title'].nunique()
unique_tv_shows = net_df[net_df['type'] == 'TV Show']['title'].nunique()

unique_films, unique_tv_shows
```

Gambar 4.10 Kode pengecekan jumlah tv dan film

Pada kode diatas kita ingin mengetahui pembagian data tentang tv dan film yang telah diterbitkan hingga sekarang. Hasilnya terdapat 6126 film dan 2664 acara tv.

```
from collections import Counter
all_genres = [genre.strip() for sublist in net_df["listed_in"].str.split(',') for genre in
sublist]
genre_counts = Counter(all_genres)
genre_df = pd.DataFrame(genre_counts.items(), columns = ['Genre' ,
'Count']).sort_values(by='Count', ascending=False)
print(genre_df)
```

Gambar 4.11 Kode pemisahan genre pada konten

Pada Gambar 4.11 dapat dilihat bahwa genre pada film dan acara TV masih berada dalam kolom yang sama dan belum terpisah. Pada tahap ini, dilakukan pemisahan genre untuk film dan acara TV, karena umumnya keduanya memiliki lebih dari satu genre. Sebagai contoh, acara TV *Blood & Water* memiliki genre *Dramas*, *Comedies*, dan *Mystery*. Berikut hasil yang diperoleh dapat dilihat di Gambar 4.12.

Genre	Count
International Movies	2752
Dramas	2427
Comedies	1674
International TV Shows	1351
Documentaries	869
Action & Adventure	859
TV Dramas	763
Independent Movies	756
Children & Family Movies	641
Romantic Movies	616
TV Comedies	581
Thrillers	577
Crime TV Shows	470
Kids' TV	451
Docuseries	395
Music & Musicals	375
Romantic TV Shows	370
Horror Movies	357

Gambar 4.12 Hasil pemisahan genre pada data film dan acara tv

```
time = net_df["year_added"] - net_df["release_year"]

# simpan data di kolom baru 'time' (ini tahun)
net_df["time"] = time
net_df.head(15)

# jalankan kode dibawah sesudah run diatas
net_df.describe(include="all")
```

Gambar 4.13 Kode untuk mengetahui durasi waktu rilis

A. Analisis Univariate

Setelah memperoleh data genre, langkah selanjutnya adalah menganalisis durasi waktu yang diperlukan Netflix untuk merilis suatu konten. Dengan analisis ini, dapat diperoleh informasi mengenai rata-rata hingga waktu maksimal yang dibutuhkan dalam proses perilsan konten tersebut.

```
time
Maksimum 93.000000
Rata-rata 2.172167
```

Gambar 4.14 Hasil waktu yang digunakan untuk merilis film dan acara tv pada data

Berdasarkan hasil analisis, rata-rata waktu rilis konten film atau acara TV adalah sekitar 2.172167 tahun, dengan rentang waktu maksimum mencapai 93 tahun. Karena rentang waktu yang diperoleh masih mencakup keseluruhan data konten secara umum, langkah selanjutnya adalah melakukan analisis lebih mendetail untuk mengetahui waktu rilis film dan acara TV. Dengan melakukan pemisahan ini, kita dapat memahami pola perilisian konten secara lebih spesifik dari segi distribusi waktu rilis.

```
net_df.loc[net_df['type'] == 'Movie', 'time2'] = net_df['year_added'] -
net_df['release_year']

net_df.loc[net_df['type'] == 'TV Show', 'time3'] = net_df['year_added'] -
net_df['release_year']

net_df.describe(include="all")
```

Gambar 4.15 Hasil waktu yang digunakan untuk merilis film dan acara tv pada data

	Maksimum	Rata-rata
time	93.0	2.172167
time2	75.0	2.502663
time3	93.0	1.493359

Gambar 4.16 Hasil waktu yang digunakan untuk merilis masing- masing film dan acara tv pada data.

Berikut hasil yang dicapai pada data:

Film: Berdasarkan hasil tersebut, dapat diketahui bahwa rata-rata waktu rilis konten film terlampaui selama 2.502663 tahun dengan nilai tahun maksimum sebesar 75 tahun.

Acara TV: Sedangkan rata-rata waktu rilis konten acara TV terlampaui selama 1.493359 tahun dengan nilai tahun maksimum sebesar 93 tahun.

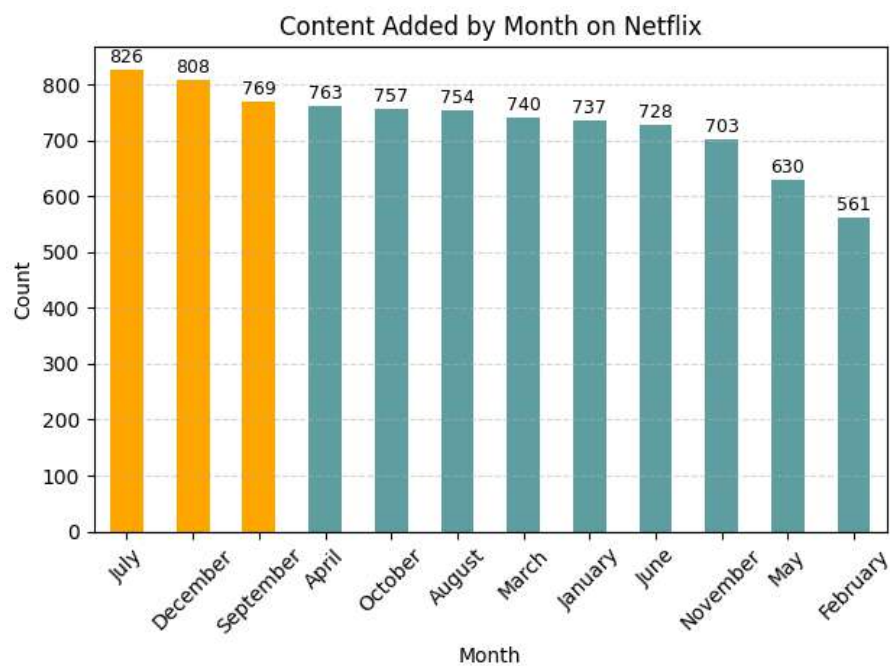
Setelah memperoleh informasi mengenai jangka waktu perilisian, langkah selanjutnya adalah menganalisis distribusi bulan perilisian film dan acara TV yang terdapat dalam data. Analisis ini bertujuan untuk mengidentifikasi pola atau tren tertentu dalam waktu perilisian, yang mungkin berkaitan dengan strategi distribusi konten yang diterapkan oleh Netflix.

```
net_df.month_name_added.value_counts()
```

Gambar 4.17 kode untuk menghitung konten per bulan

month_name_added	count
July	827
December	812
September	769
April	763
October	760
August	754
March	741
January	737
June	728
November	705
May	632
February	562

Gambar 4.18 Jumlah perilsan film dan acara tv pada setiap bulan



Gambar 4.19 Visualisasi Hasil Jumlah perilsan film dan acara tv pada setiap bulan

B. Analisis Bivariate

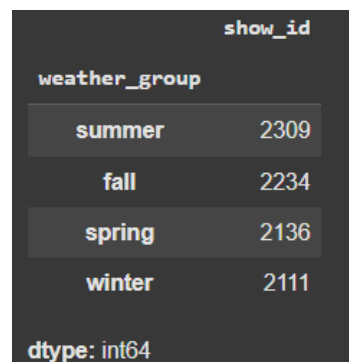
Hasil pada Gambar 4.18 sudah memberikan gambaran awal mengenai pola algoritma yang digunakan Netflix dalam perilisian film dan acara TV. Namun, untuk memperoleh pemahaman yang lebih jelas dan terstruktur, akan dilakukan pengelompokkan nama bulan ke dalam empat kategori musim.

```
net_df["weather_group"] = net_df["month_added"].apply(
    lambda x: "winter" if x <= 2 or x == 12
    else "spring" if 3 <= x <= 5
    else "summer" if 6 <= x <= 8
    else "fall"
)

season_counts =
net_df.groupby(by="weather_group")["show_id"].nunique().sort_values(ascending=False)

season_counts
```

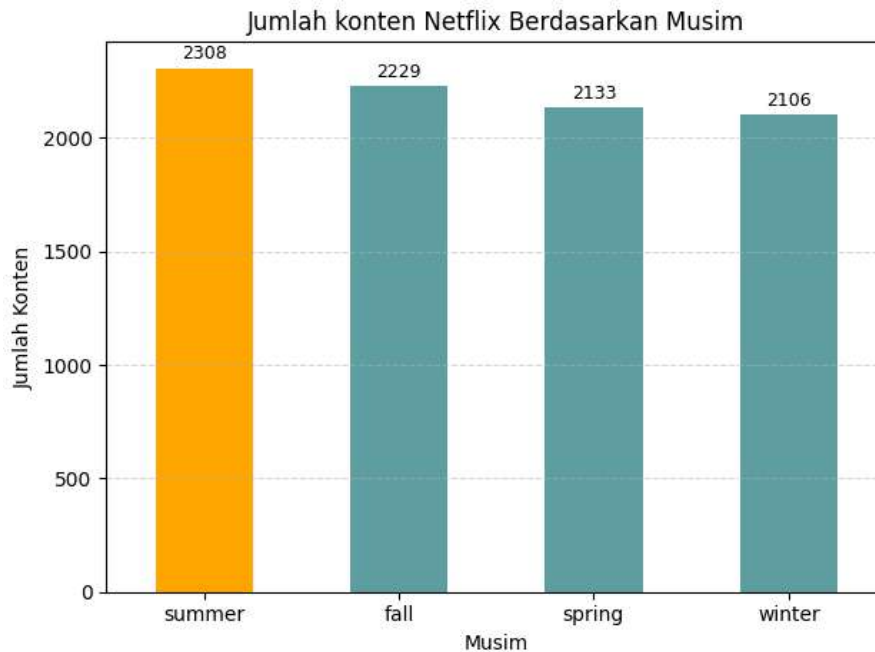
Gambar 4.20 kode untuk mengubah bulan ke 4 musim



weather_group	show_id
summer	2309
fall	2234
spring	2136
winter	2111

dtype: int64

Gambar 4.21 Hasil perilisian sesuai musim



Gambar 4.22 Hasil visualisasi perilisan sesuai musim

Berikut adalah struktur pembagian musim berdasarkan tanggal: *Fall* (23 September – 20 Desember), *Winter* (21 Desember – 20 Maret), *Spring* (21 Maret – 20 Juni), dan *Summer* (21 Juni – 22 September). Berdasarkan pembagian musim tersebut, dapat disimpulkan bahwa Netflix lebih memilih untuk merilis film dan acara TV pada musim panas (*Summer*), yang kemungkinan besar dikaitkan dengan tingginya tingkat konsumsi hiburan di periode tersebut.

Jika kita ingin mengetahui perilisan suatu film atau acara tv paling banyak di bulan apa kita bisa menggunakan kode berikut:

```
tv_shows_data = net_df[net_df['type'] == 'TV Show'].copy()
movies_data = net_df[net_df['type'] == 'Movie'].copy()

movies_data['month'] = movies_data['date_added'].dt.month
tv_shows_data['month'] = tv_shows_data['date_added'].dt.month

movies_monthly_count = movies_data.groupby('month')['title'].nunique().reset_index()

tv_shows_monthly_count =
tv_shows_data.groupby('month')['title'].nunique().reset_index()

best_month_movies =
movies_monthly_count.sort_values('title', ascending=False).iloc[0]
```

```

best_month_tv_shows =
tv_shows_monthly_count.sort_values('title', ascending=False).iloc[0]

best_month_movies, best_month_tv_shows

#bulan yang sering upload

```

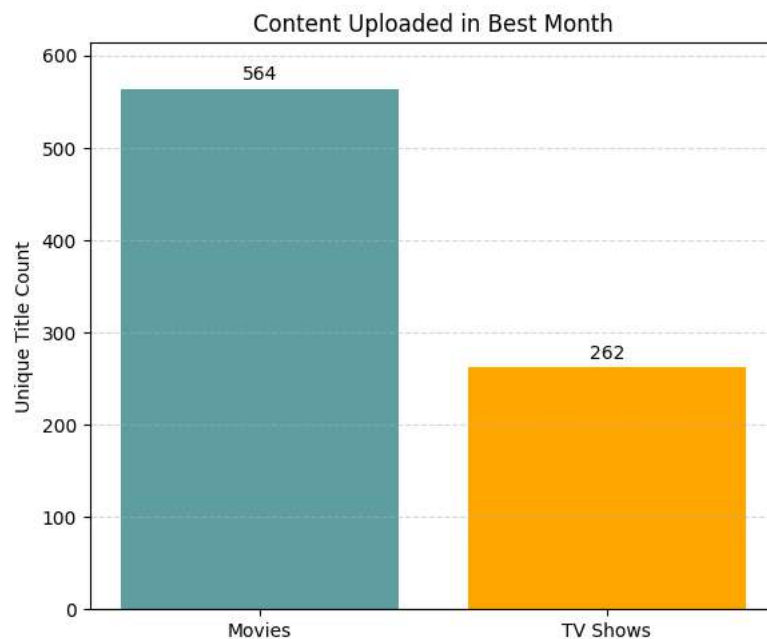
Gambar 4.23 kode untuk menemukan aktivitas konten yang sering dirilis pada suatu bulan

```

(month      7
 title     565
 Name: 6, dtype: int64,
 month     12
 title     265
 Name: 11, dtype: int64)

```

Gambar 4.24 Bulan paling dominan rilisnya film/acara tv



Gambar 4.25 Hasil visualisasi bulan juli sebagai dominan rilisnya film dan acara tv

Pada Gambar 4.24, hasil analisis menunjukkan bahwa film paling banyak dirilis pada bulan Juli, dengan jumlah mencapai 565 rilis (*Summer*), sementara acara TV paling banyak dirilis pada bulan Desember, dengan jumlah mencapai 265 rilis (*Winter*).

Dalam perilsan film dan acara TV, penting untuk memperhatikan kesesuaian usia penonton. Pada Gambar 4.25, ditampilkan distribusi konten berdasarkan kategori usia untuk film dan acara TV yang terdapat dalam data Netflix. Analisis ini bertujuan untuk memahami bagaimana Netflix mengelompokkan tayangan berdasarkan target usia audiensnya.

```
unique_titles_per_age_ratings =
unnested_age_ratings.groupby('age_ratings')['show_id'].nunique()

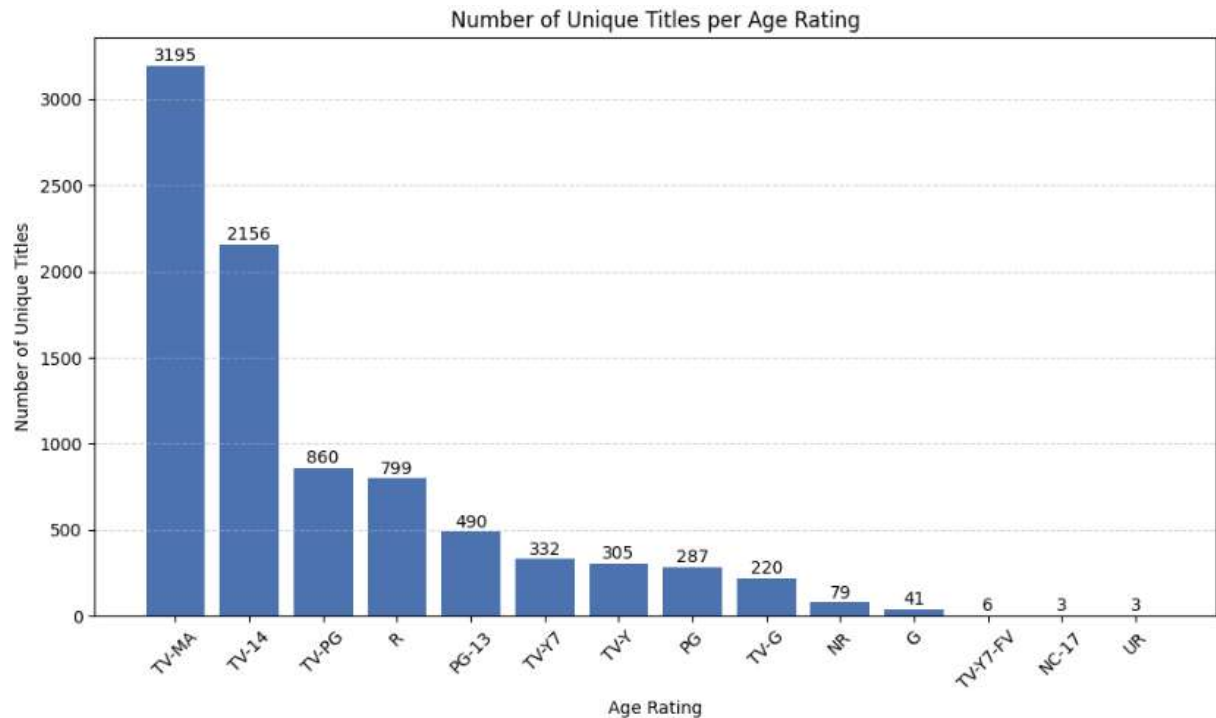
# Sorting the result in descending order
unique_titles_per_age_ratings_sorted =
unique_titles_per_age_ratings.sort_values(ascending=False)

# Displaying the result
unique_titles_per_age_ratings_sorted
```

Gambar 4.26 kode untuk menemukan jumlah klasifikasi usia pada konten

age_ratings	show_id
TV-MA	3205
TV-14	2157
TV-PG	861
R	799
PG-13	490
TV-Y7	333
TV-Y	306
PG	287
TV-G	220

Gambar 4.27 Jumlah klasifikasi usia film/acara tv sesuai usia penonton



Gambar 4.28 Hasil Visualisasi jumlah klasifikasi usia film dan acara tv sesuai usia penonton

Analisis tidak berhenti sampai di situ saja. Untuk mendapatkan pemahaman yang lebih komprehensif, diperlukan juga penarikan kesimpulan terkait pola atau algoritma perilsan film dan acara TV di Netflix berdasarkan distribusi bulan rilis. Dengan demikian, dapat diidentifikasi kecenderungan waktu tertentu yang dipilih oleh Netflix untuk merilis kontennya, yang kemungkinan besar berkaitan dengan strategi menarik perhatian audiens pada periode tertentu.

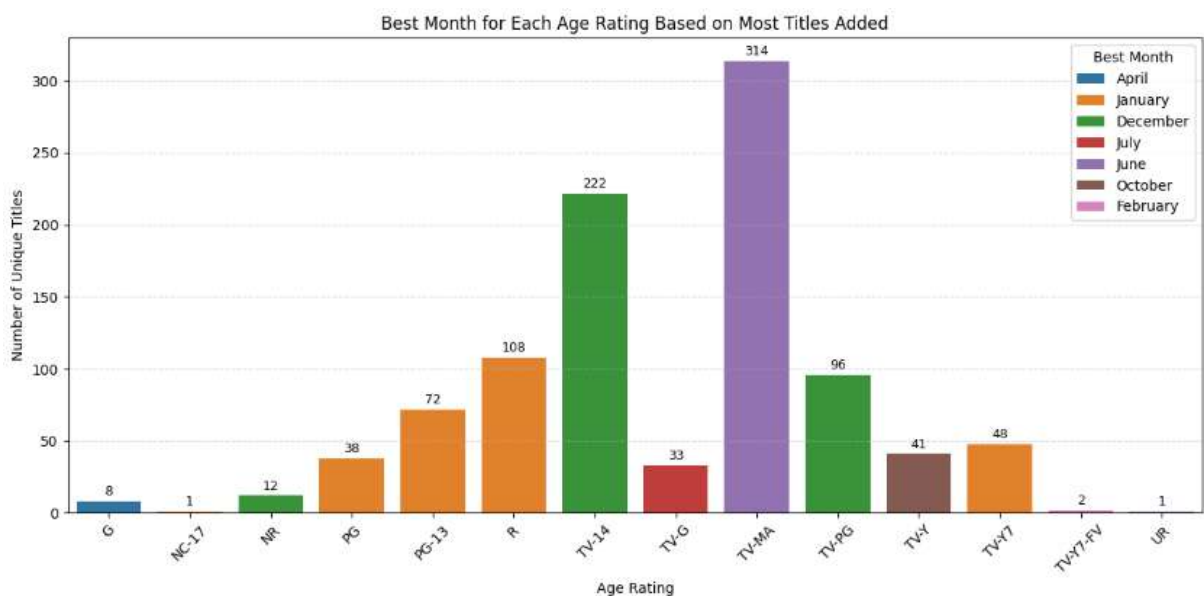
```
net_df['month'] = net_df['date_added'].dt.month
shows_monthly_rate = net_df.groupby(['age_ratings',
'month'])['title'].nunique().reset_index()

best_months_by_rating = shows_monthly_rate.sort_values(['age_ratings', 'title'],
ascending=[True, False]).drop_duplicates('age_ratings')
print(best_months_by_rating)
```

Gambar 4.29 kode untuk menemukan jumlah klasifikasi usia konten pada bulan

age_ratings	month	title
3	G	4
11	NC-17	1
25	NR	12
26	PG	1
38	PG-13	1
50	R	1
73	TV-14	12
80	TV-G	7
91	TV-MA	6
109	TV-PG	12
119	TV-Y	10
122	TV-Y7	1
134	TV-Y7-FV	2
138	UR	6

Gambar 4.30 Perilisan terbanyak klasifikasi usia film/acara tv pada bulan.



Gambar 4.31 Hasil Visualisasi perilisan terbanyak klasifikasi usia film dan acara tv pada bulan.

Dari Gambar 4.31, dapat ditarik beberapa kesimpulan, salah satunya adalah ditemukannya fakta bahwa konten dengan kategori *age ratings* yang aman untuk anak-anak atau yang dapat ditonton bersama keluarga cenderung lebih banyak dirilis pada musim liburan. Pola ini menunjukkan bahwa Netflix kemungkinan sengaja memanfaatkan momentum liburan untuk menghadirkan konten yang dapat dinikmati seluruh anggota keluarga secara bersama-sama.

Selain mengklasifikasikan berdasarkan kategori usia, penting juga untuk menganalisis asal negara yang merilis film dan acara TV. Informasi ini diperlukan untuk memahami sebaran produksi konten di Netflix, serta untuk mengidentifikasi negara-negara yang berkontribusi besar dalam penyediaan konten pada platform tersebut:

```
all_country = [country.strip() for sublist in net_df["country"].str.split(',') for country in
sublist]

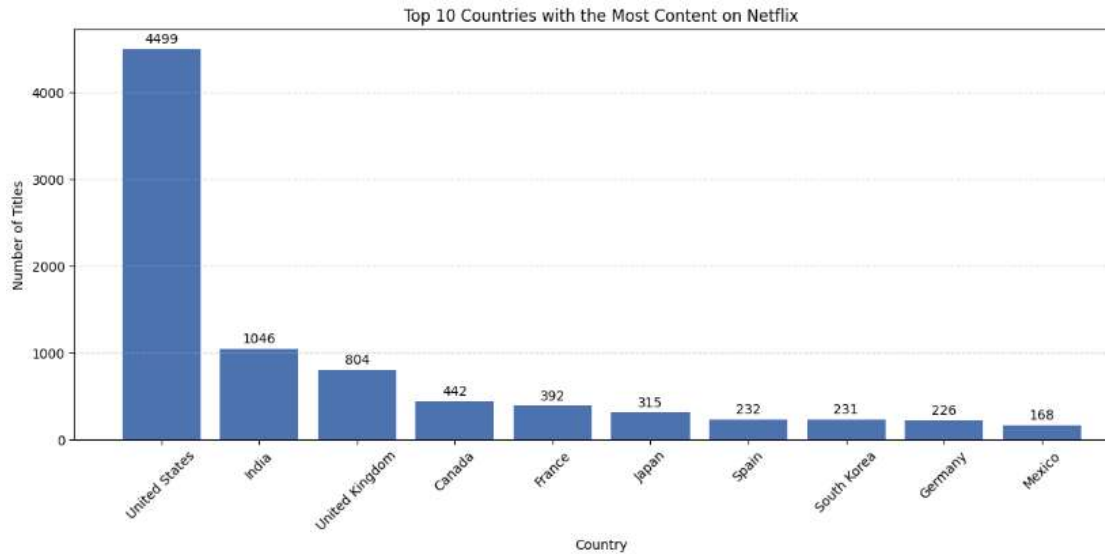
country_count = Counter(all_country)
country_df = pd.DataFrame(country_count.items(), columns=['country',
'Count']).sort_values(by='Count', ascending=False)

print(country_df)
```

Gambar 4.32 Kode untuk menemukan total jumlah konten pada suatu negara

country	Count
United States	4499
India	1046
United Kingdom	804
Canada	442
France	392
...	...
Sudan	1
Panama	1
Uganda	1
East Germany	1
Montenegro	1

Gambar 4.33 Jumlah negara dalam perilsan film/acara tv



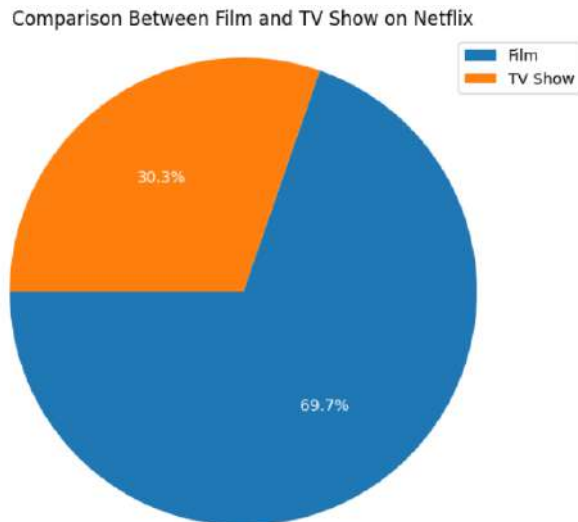
Gambar 4.34 Jumlah negara dalam perilsan film/acara tv

Dari hasil analisis, ditemukan bahwa Amerika Serikat (USA) merupakan negara yang paling banyak merilis film dan acara TV di Netflix. Hal ini dapat dimengerti mengingat penggunaan bahasa Inggris sebagai bahasa internasional, yang memungkinkan jangkauan konten menjadi lebih luas dan lebih mudah diterima oleh audiens global.

4.4 *Visualize and interpret the result*

A. Profiling data pengguna

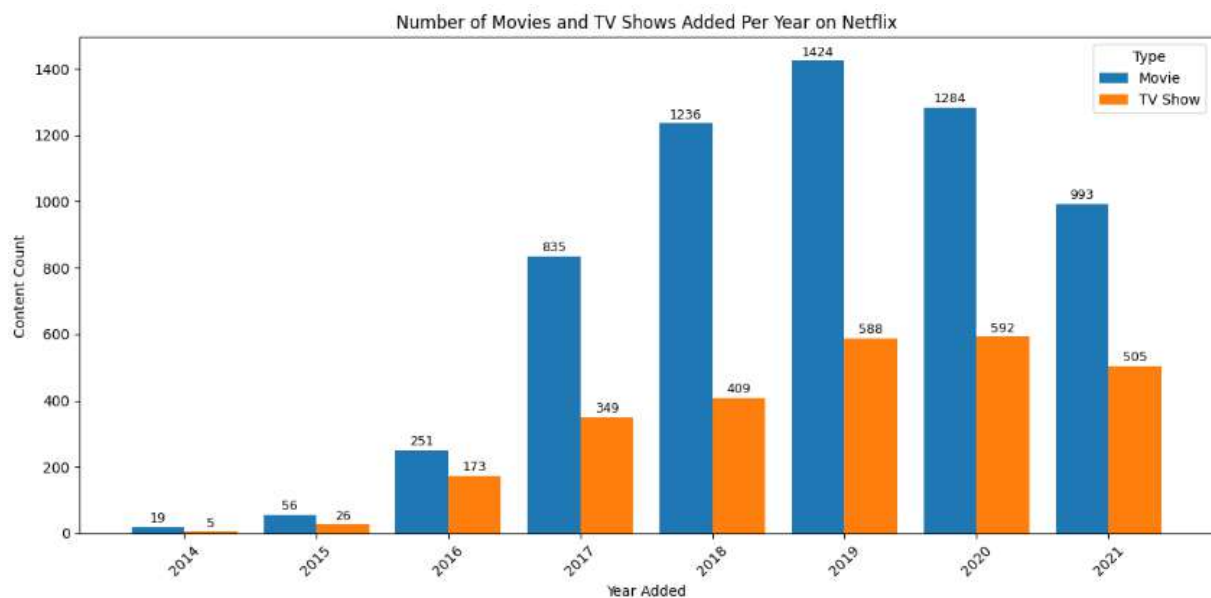
Dalam kerangka penelitian ini, Analisis dengan metode kmeans diawali dengan visualisasi data akan diawali penyajian melalui pemahaman umum atau mendasar terlebih dulu agar dapat memberikan wawasan tentang berbagai aspek platform Netflix berdasarkan analisis data film dan acara TV yang dibuat oleh berbagai negara di dunia.



Gambar 4.35 Visualisasi perbandingan film dan acara tv

Pada visualisasi diatas film memiliki persentasi 69,7% dan acara tv memiliki persentase 30.3% ini cocok seperti hasil yang kita dapatkan sebelumnya sejumlah 6126 film dan 2664 acara tv.

Selain pembagian film dan acara tv, kita juga perlu mendapatkan informasi tentang tahun rilisnya dari kategori tersebut.



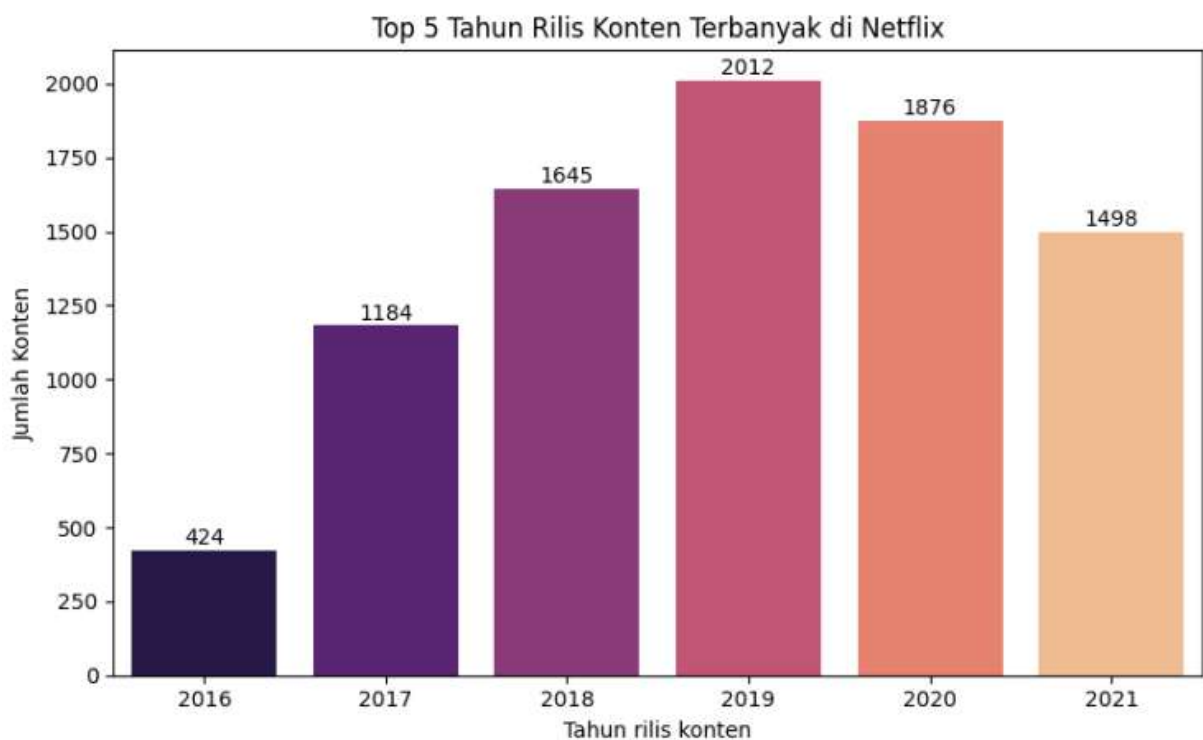
Gambar 4.36 Visualisasi perbandingan film dan acara tv per tahun

Dalam memahami tren pertumbuhan konten pada platform Netflix, dilakukan analisis terhadap jumlah film dan acara TV yang ditambahkan setiap tahunnya, dimulai dari tahun 2014

hingga 2021. Visualisasi ini berguna untuk mengamati pola ekspansi dan perubahan strategi distribusi konten yang dijalankan oleh Netflix selama periode tersebut.

Berdasarkan grafik tersebut, dapat dilihat bahwa sejak tahun 2016 terjadi lonjakan signifikan dalam jumlah konten yang ditambahkan ke Netflix. Penambahan konten terus meningkat dari tahun ke tahun, hingga mencapai puncaknya pada tahun 2019 dengan total 1.424 film dan 588 acara TV. Ini menjadi indikasi kuat bahwa tahun 2019 merupakan periode ekspansi besar-besaran bagi Netflix dalam memperkaya kontennya secara global.

Meskipun tahun 2020 dan 2021 menunjukkan penurunan jumlah konten yang ditambahkan, angka tersebut masih tergolong tinggi. Pada tahun 2020, terdapat 1.284 film dan 592 acara TV, sedangkan pada tahun 2021 terdapat 993 film dan 505 acara TV.



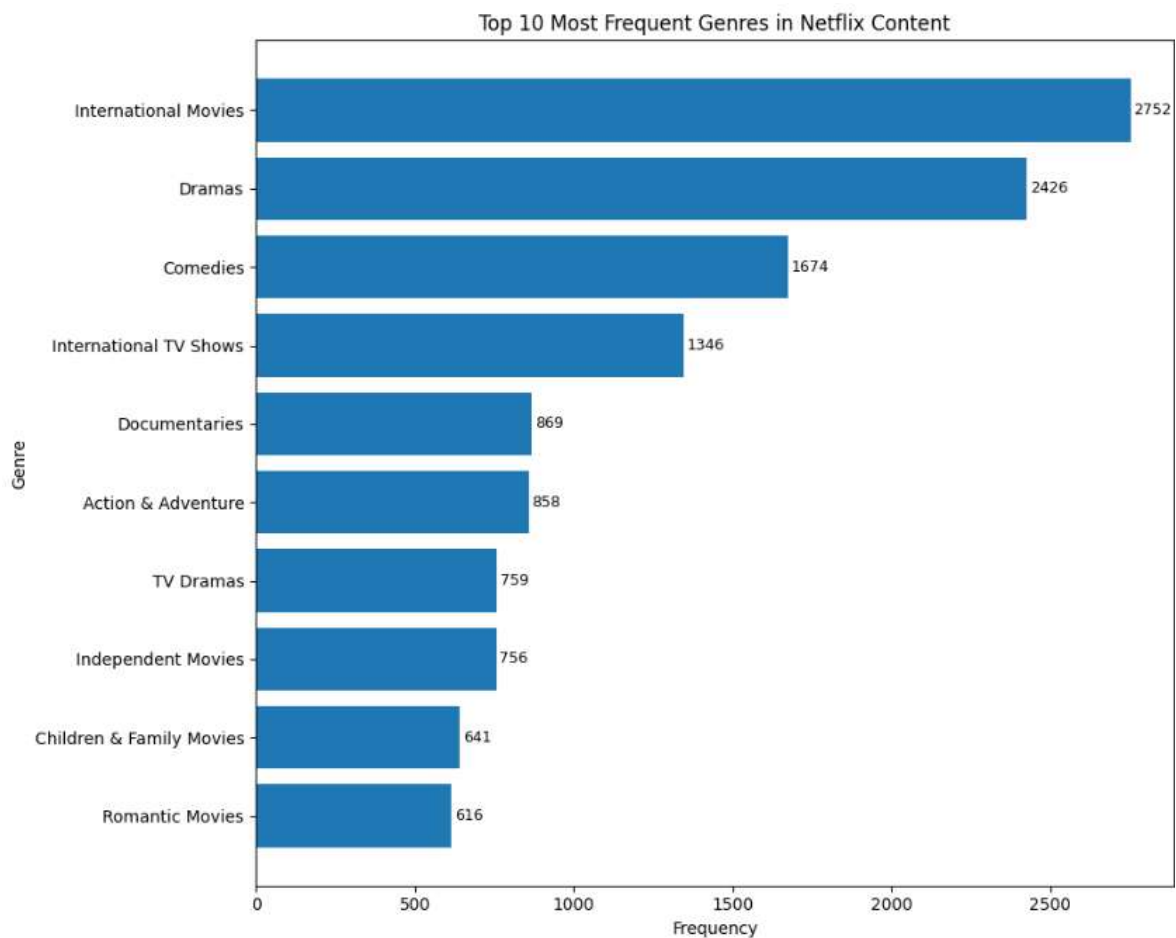
Gambar 4.37 Visualisasi top 5 tahun rilis konten per tahun

Gambar 4.37 menunjukkan lima tahun dengan jumlah konten rilis terbanyak di platform Netflix, yaitu dari tahun 2016 hingga 2021. Berdasarkan visualisasi tersebut, dapat dilihat bahwa tahun 2019 menjadi puncak tertinggi dengan total 2.012 konten yang dirilis, diikuti oleh tahun 2020 (1.876 konten) dan tahun 2018 (1.645 konten). Sementara itu, meskipun tahun 2021

masih mencatatkan angka yang cukup tinggi yaitu 1.498 konten, terdapat penurunan dibandingkan tahun sebelumnya. Tahun 2016 menjadi titik awal dari tren peningkatan signifikan, dengan jumlah konten hanya 424 pada tahun tersebut.

Pola ini menunjukkan bahwa dalam kurun waktu 2016 hingga 2020, Netflix mengalami ekspansi besar-besaran dalam produksi dan perilisian konten global. Hal ini dapat dikaitkan dengan strategi Netflix dalam memperluas pasar internasional. Secara keseluruhan, grafik ini memberikan wawasan tentang bagaimana Netflix secara agresif meningkatkan jumlah rilis kontennya dalam beberapa tahun terakhir, dengan puncak aktivitas terjadi menjelang pandemi.

Genre juga merupakan aspek penting yang perlu diperhatikan, karena dengan menganalisis genre suatu film atau acara TV, kita dapat memperoleh gambaran yang lebih jelas mengenai algoritma yang digunakan oleh Netflix dalam menentukan dan menyarankan konten kepada pengguna berdasarkan preferensi genre tertentu.

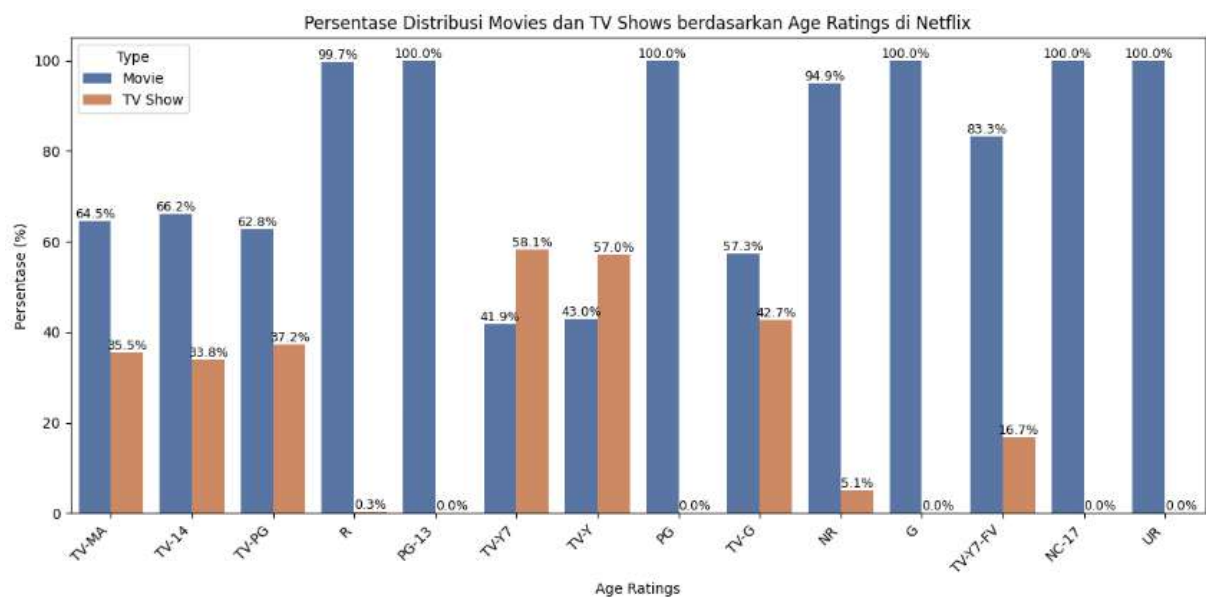


Gambar 4.38 Visualisasi genre film dan acara tv Netflix pada data

Berdasarkan hasil analisis, International Movies menempati posisi teratas dengan total sebanyak 2.752 judul, diikuti oleh Dramas sebanyak 2.426 judul dan Comedies sebanyak 1.674 judul. Genre-genre ini menunjukkan bahwa Netflix mengutamakan konten yang bersifat global dan memiliki daya tarik emosional atau hiburan ringan yang dapat dinikmati oleh berbagai lapisan audiens.

Genre International TV Shows juga cukup dominan dengan 1.346 judul, menandakan ekspansi Netflix dalam memperluas distribusi serial dari berbagai negara. Genre dokumenter dan aksi juga hadir dalam jumlah yang signifikan, masing-masing sebanyak 869 dan 858 judul, mengindikasikan bahwa konten informatif dan penuh ketegangan tetap memiliki tempat tersendiri di hati penonton.

Lalu bagaimana perbandingannya jika perilsan konten berdasarkan klasifikasi usia tertentu ?



Gambar 4.39 Visualisasi perbandingan konten klasifikasi usia tertentu

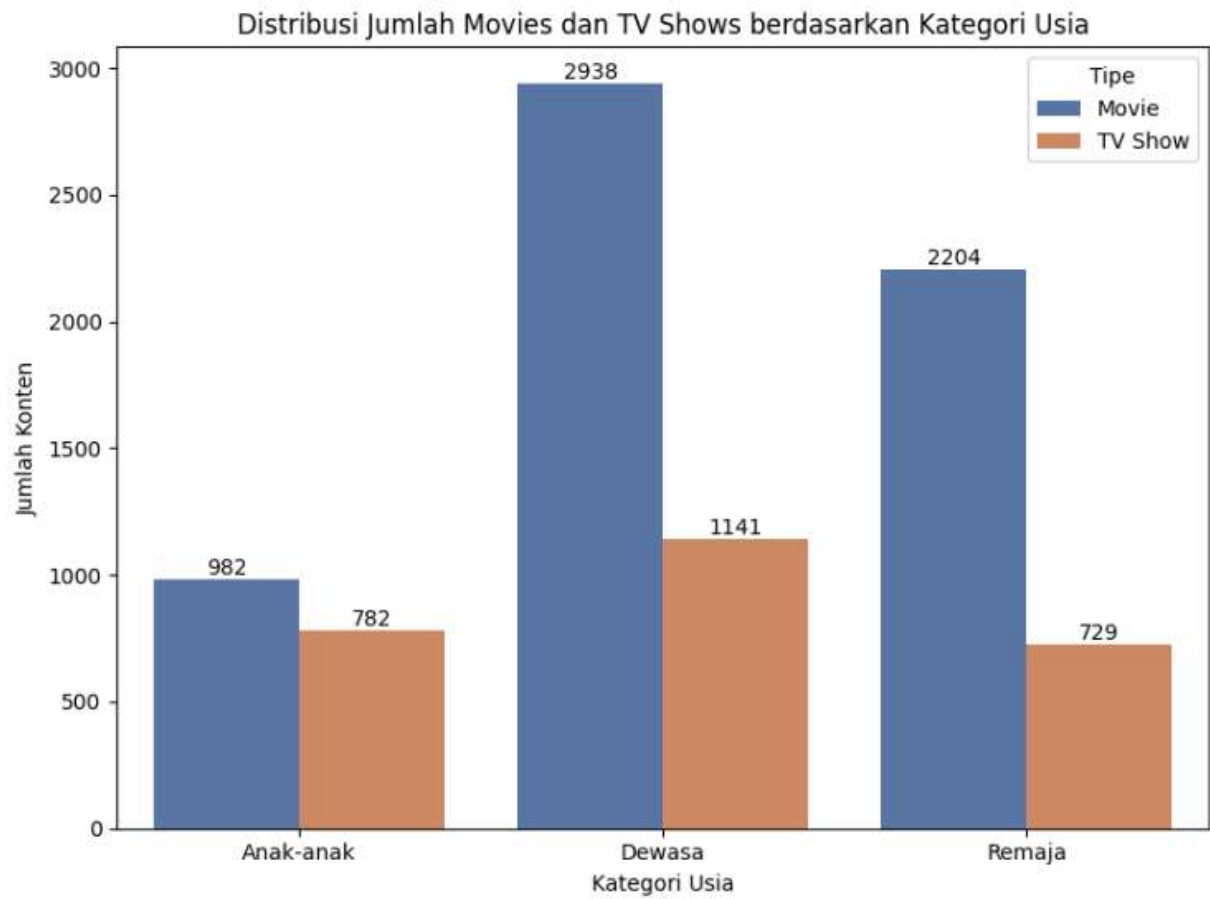
Gambar 4.39 memperlihatkan distribusi persentase konten Netflix berdasarkan klasifikasi usia (age ratings) dan jenis tayangan (Movie dan TV Show). Dari grafik tersebut, terlihat bahwa kategori usia TV-MA, TV-14, dan TV-PG merupakan tiga klasifikasi usia dengan jumlah

konten terbanyak, baik pada film maupun acara TV. Namun, persentase dominasi tayangan film sangat jelas pada hampir semua kategori klasifikasi usia.

Misalnya, klasifikasi R, PG-13, PG, dan G hampir seluruhnya didominasi oleh film, bahkan pada beberapa kategori seperti PG-13, PG, dan G, proporsi TV Show adalah 0%. Sebaliknya, kategori TV-Y7, TV-Y, dan TV-Y7-FV memiliki proporsi TV Show yang lebih tinggi dibandingkan klasifikasi usia lainnya, karena segmen anak-anak lebih sering dikembangkan dalam format serial. Hal ini mengindikasikan bahwa Netflix lebih sering merilis konten ramah anak dalam bentuk TV Show, sedangkan konten untuk remaja dan dewasa lebih banyak diproduksi dalam bentuk film. Pola ini menunjukkan adanya segmentasi produksi berdasarkan preferensi usia penonton dan format tayangan yang sesuai.

Untuk memahami bagaimana Netflix menargetkan segmen audiensnya, dilakukan pengelompokan konten berdasarkan kategori usia. Pengelompokan ini penting untuk menganalisis arah strategi distribusi konten, khususnya dalam hal kesesuaian tayangan terhadap kelompok usia tertentu, serta mengevaluasi keberagaman dan keseimbangan konten yang ditawarkan kepada penonton dari berbagai rentang umur.

Gambar di bawah ini memperlihatkan distribusi jumlah film dan acara TV berdasarkan tiga kelompok usia, yaitu anak-anak, remaja, dan dewasa:



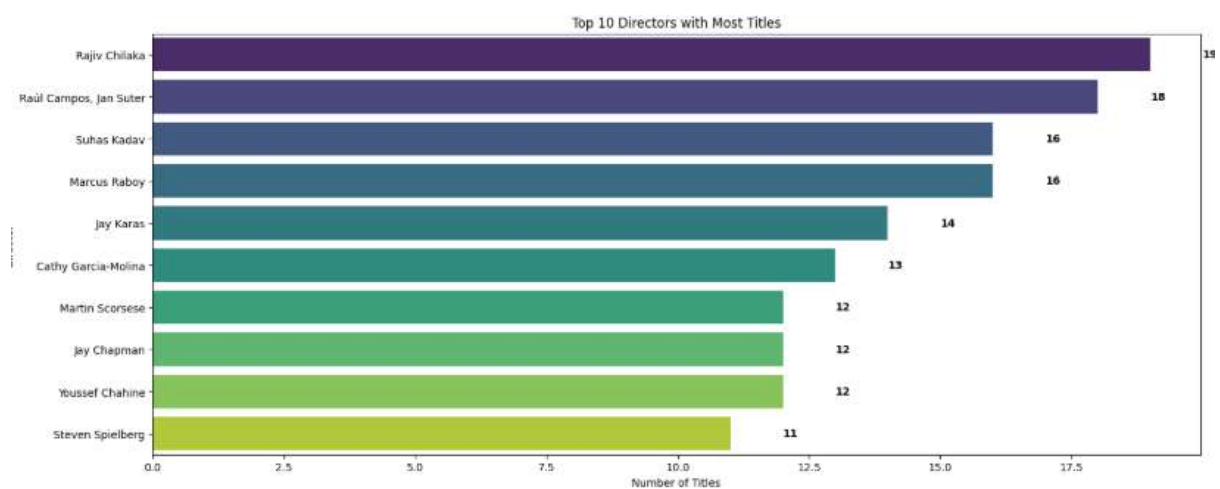
Gambar 4.40 Visualisasi Distribusi perbandingan konten klasifikasi usia

Berdasarkan visualisasi tersebut, terlihat bahwa konten dewasa mendominasi baik pada format film (2.938 judul) maupun acara TV (1.141 judul). Hal ini menunjukkan bahwa Netflix memiliki fokus yang cukup kuat dalam menyediakan konten untuk penonton usia dewasa. Sementara itu, konten untuk anak-anak dan remaja juga tersedia dalam jumlah signifikan, terutama pada format film (masing-masing 982 dan 2.204 judul), namun jumlahnya lebih rendah dibandingkan konten dewasa. Pada format TV Show, konten anak-anak (782 judul) dan remaja (729 judul) relatif seimbang dan mencerminkan pendekatan Netflix dalam menghadirkan tayangan berseri yang ramah keluarga dan sesuai usia.

Temuan ini memperlihatkan bahwa meskipun Netflix menawarkan variasi kategori usia, terdapat kecenderungan yang kuat untuk menghadirkan lebih banyak konten yang ditujukan bagi penonton dewasa, baik dalam bentuk film maupun acara TV. Strategi ini kemungkinan besar berhubungan dengan segmentasi pasar utama Netflix yang didominasi oleh pengguna

usia remaja akhir hingga dewasa muda, serta tingginya konsumsi konten dewasa dalam industri hiburan digital secara global.

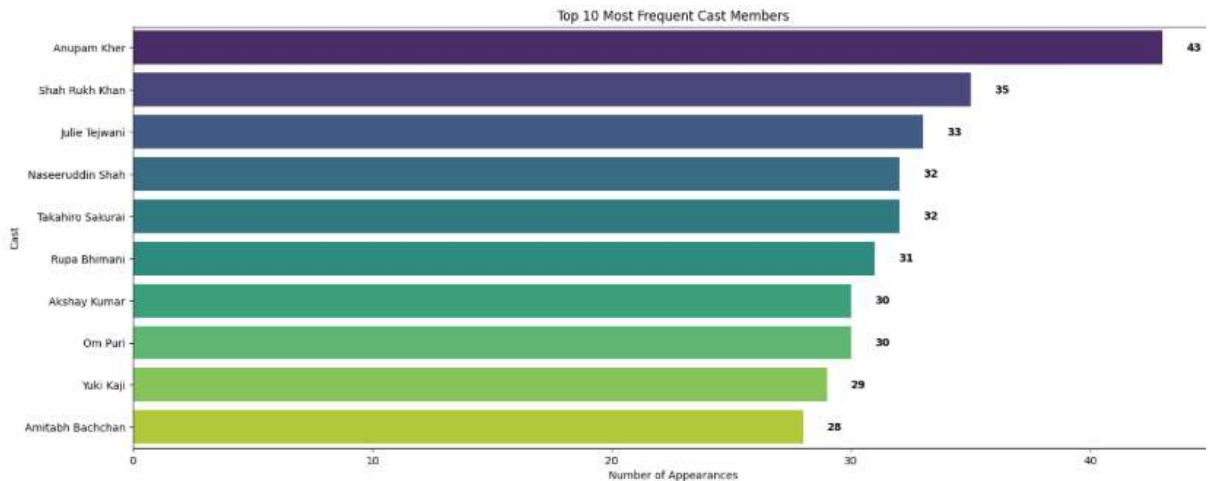
Pada Gambar 4.41 di bawah ditampilkan visualisasi sepuluh besar sutradara dengan jumlah judul terbanyak yang tayang di platform Netflix. Grafik ini disusun berdasarkan frekuensi kemunculan nama sutradara dalam dataset, yang menunjukkan kontribusi mereka dalam menyediakan konten bagi platform tersebut.



Gambar 4.41 Visualisasi Top 10 Directors

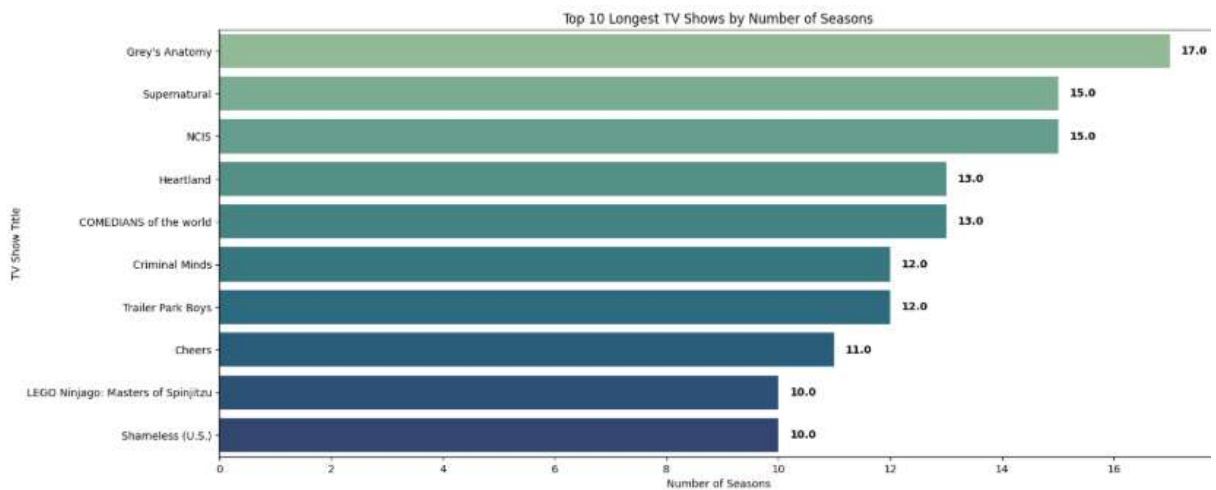
Berdasarkan visualisasi di atas, Rajiv Chilaka menempati posisi teratas dengan total 19 judul, disusul oleh Raúl Campos dan Jan Suter sebanyak 18 judul, serta Suhas Kadav dan Marcus Raboy masing-masing dengan 16 judul. Sutradara terkenal seperti Martin Scorsese dan Steven Spielberg juga muncul dalam daftar, masing-masing dengan 12 dan 11 judul.

Netflix tidak hanya menampilkan karya dari sutradara ternama internasional, tetapi juga memberi ruang cukup besar bagi kreator dari berbagai wilayah, termasuk dari India dan Amerika Latin, untuk memproduksi konten dalam jumlah besar. Hal ini mencerminkan diversifikasi konten dan strategi global Netflix dalam menjangkau audiens dari berbagai latar belakang budaya.



Gambar 4.42 Visualisasi Top 10 Cast Members

Dari grafik tersebut Anupam Kher menjadi aktor dengan kemunculan terbanyak, yaitu sebanyak 43 kali, disusul oleh Shah Rukh Khan sebanyak 35 kali, dan Julie Tejwani sebanyak 33 kali. Aktor-aktor lain yang mendominasi daftar ini antara lain Naseeruddin Shah, Takahiro Sakurai, serta Amitabh Bachchan. Mayoritas dari daftar tersebut berasal dari industri film India, yang menunjukkan bahwa konten Bollywood memainkan peran penting dalam katalog Netflix, khususnya dalam menjangkau audiens Asia maupun diaspora global. Selain itu, kehadiran aktor Jepang seperti Takahiro Sakurai dan aktor pengisi suara seperti Yuki Kaji juga menandakan tingginya frekuensi konten Jepang.

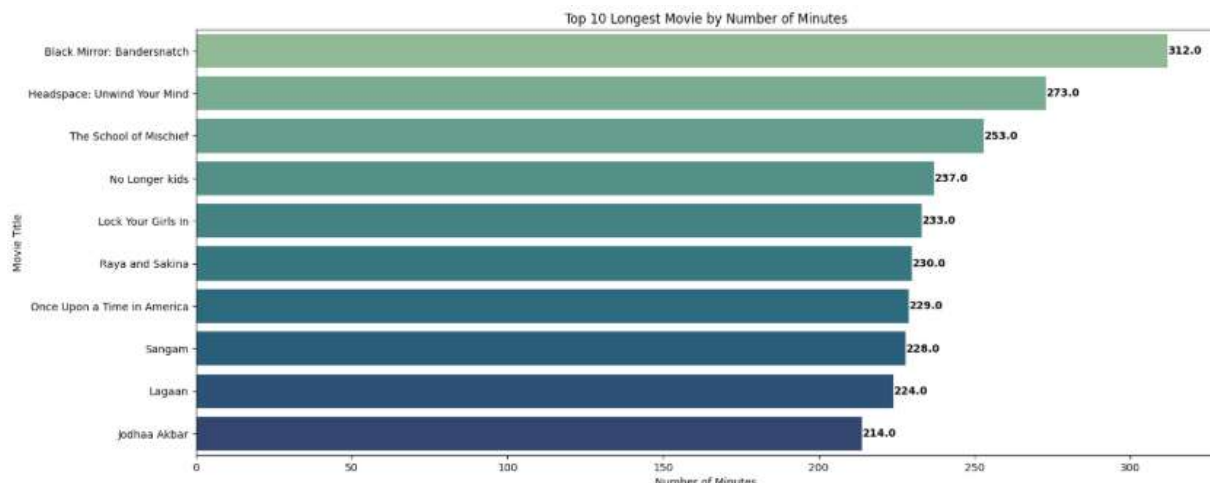


Gambar 4.43 Visualisasi Top 10 Durasi TV Shows

Grey's Anatomy menempati posisi teratas dengan total 17 musim, disusul oleh *Supernatural* dan *NCIS*, masing-masing dengan 15 musim. Judul lain seperti *Heartland*, *COMEDIANS of the*

world, dan *Criminal Minds* juga termasuk dalam daftar dengan lebih dari 12 musim. TV show dengan musim yang panjang biasanya mencerminkan popularitas yang konsisten, serta loyalitas penonton yang tinggi. Produksi jangka panjang ini juga menunjukkan bahwa Netflix tidak hanya menyediakan konten baru, tetapi juga mempertahankan tayangan klasik atau legendaris yang telah memiliki basis penggemar luas.

Selain itu, dominasi judul-judul dari Amerika Serikat seperti *NCIS*, *Criminal Minds*, dan *Cheers* menunjukkan bahwa konten TV dari Amerika masih menjadi andalan dalam koleksi tayangan berseri Netflix. Keberadaan judul seperti *LEGO Ninjago* juga menandakan keikutsertaan konten animasi dan keluarga dalam tren ini.

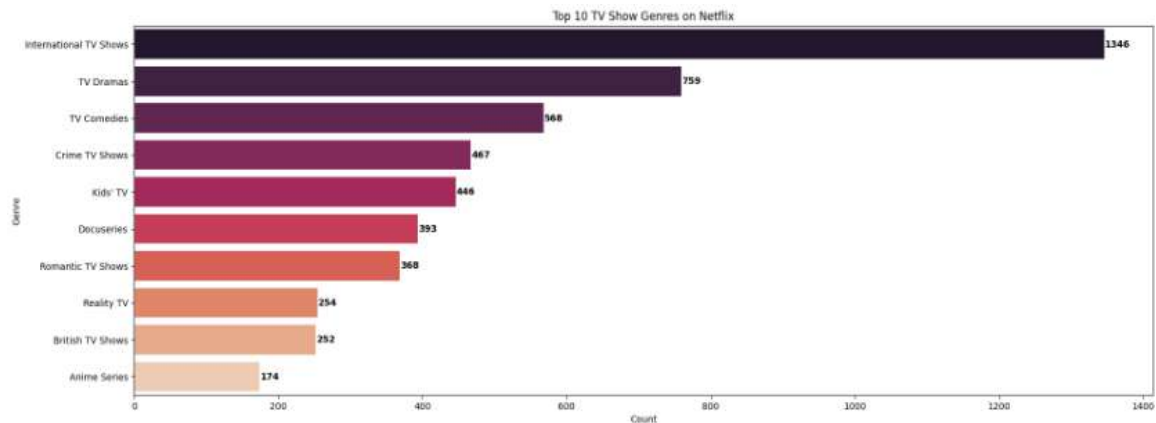


Gambar 4.44 Visualisasi Top 10 Durasi Movies

Film "Black Mirror: Bandersnatch" menempati posisi teratas dengan durasi mencapai 312 menit, disusul oleh "Headspace: Unwind Your Mind" (273 menit), dan "The School of Mischief" (253 menit). Judul-judul lainnya memiliki durasi di atas 210 menit, yang tergolong sangat panjang untuk kategori film.

Film berdurasi panjang umumnya menawarkan alur cerita yang kompleks, karakter yang mendalam, atau bahkan interaktivitas, seperti halnya *Black Mirror: Bandersnatch* yang mengusung konsep film interaktif, di mana pengguna memilih jalannya cerita. Menariknya, beberapa film dalam daftar seperti *Lagaan*, *Sangam*, dan *Jodhaa Akbar* berasal dari India, yang

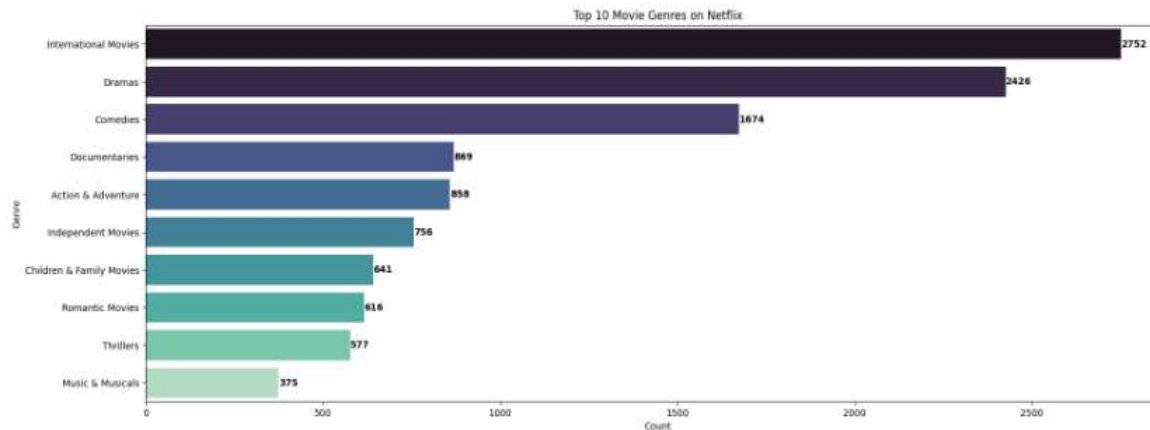
menandakan bahwa film Bollywood sering kali menghadirkan cerita dengan durasi yang lebih panjang dari standar film Barat.



Gambar 4.45 Visualisasi Top 10 Genres TV Shows

Genre International TV Shows mendominasi daftar dengan 1.346 judul, yang menunjukkan bahwa Netflix secara aktif mengkurasi dan mendistribusikan konten dari berbagai negara untuk menjangkau audiens global. Ini juga mendukung strategi ekspansi internasional Netflix yang bertujuan merangkul preferensi lokal dan memperluas basis pelanggannya di berbagai wilayah. Genre berikutnya yang paling banyak tersedia adalah TV Dramas (759 judul), TV Comedies (568 judul), dan Crime TV Shows (467 judul). Dominasi genre drama dan komedi menunjukkan minat tinggi dari pengguna terhadap cerita emosional dan hiburan ringan.

Keberadaan genre seperti Kids' TV (446 judul), DocuSeries (393 judul), dan Reality TV (254 judul) menunjukkan bahwa Netflix juga menargetkan segmen penonton yang beragam, termasuk anak-anak, penikmat dokumenter, hingga pemirsa yang menyukai konten *slice of life*.

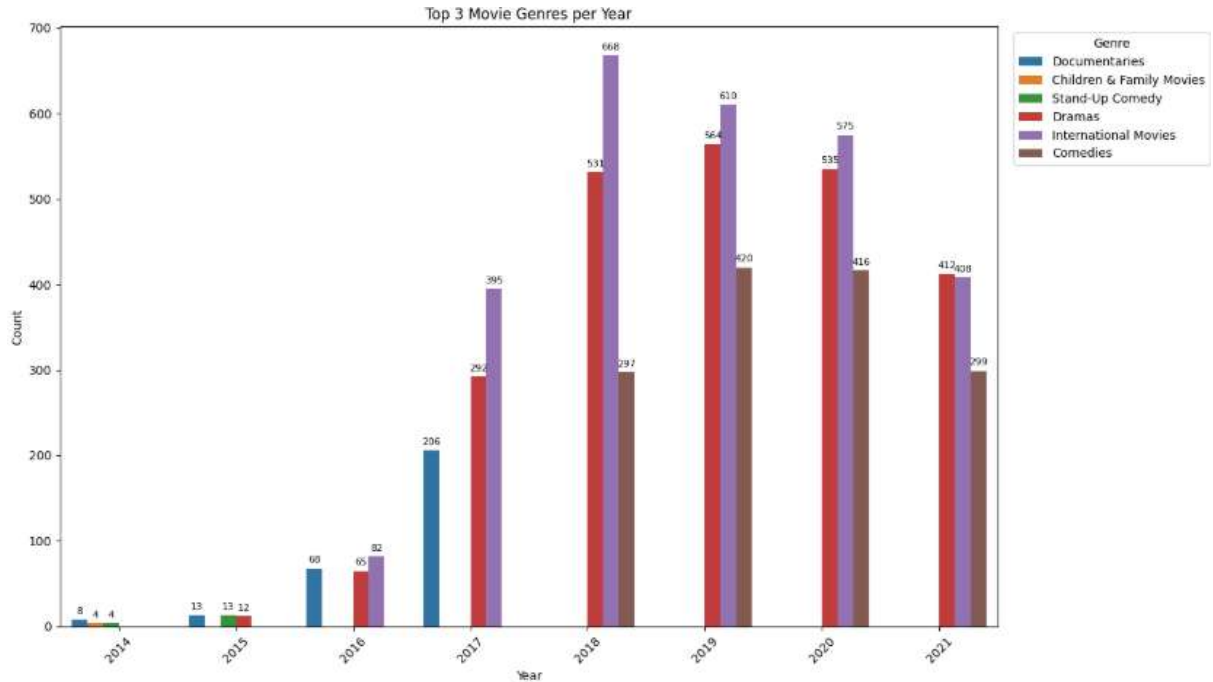


Gambar 4.46 Visualisasi Top 10 Genres Movies

Berdasarkan visualisasi, genre International Movies berada di peringkat teratas dengan 2.752 judul, menunjukkan bahwa Netflix secara aktif memperluas jangkauan kontennya secara global, serta mendukung beragam budaya dan bahasa melalui distribusi film dari berbagai negara. Genre Dramas menyusul dengan 2.426 judul, yang menunjukkan bahwa cerita dengan narasi mendalam dan emosional masih menjadi salah satu daya tarik utama bagi penonton.

Sementara itu, Comedies berada di posisi ketiga dengan 1.674 judul, menandakan konten ringan dan menghibur dalam platform ini cukup populer. Genre lain yang juga banyak tersedia meliputi Documentaries (869 judul), Action & Adventure (858 judul), dan Independent Movies (756 judul), menunjukkan bahwa Netflix menyediakan variasi genre untuk menjangkau berbagai jenis penonton, mulai dari yang menginginkan informasi faktual, aksi yang memacu adrenalin, hingga karya sinema independen.

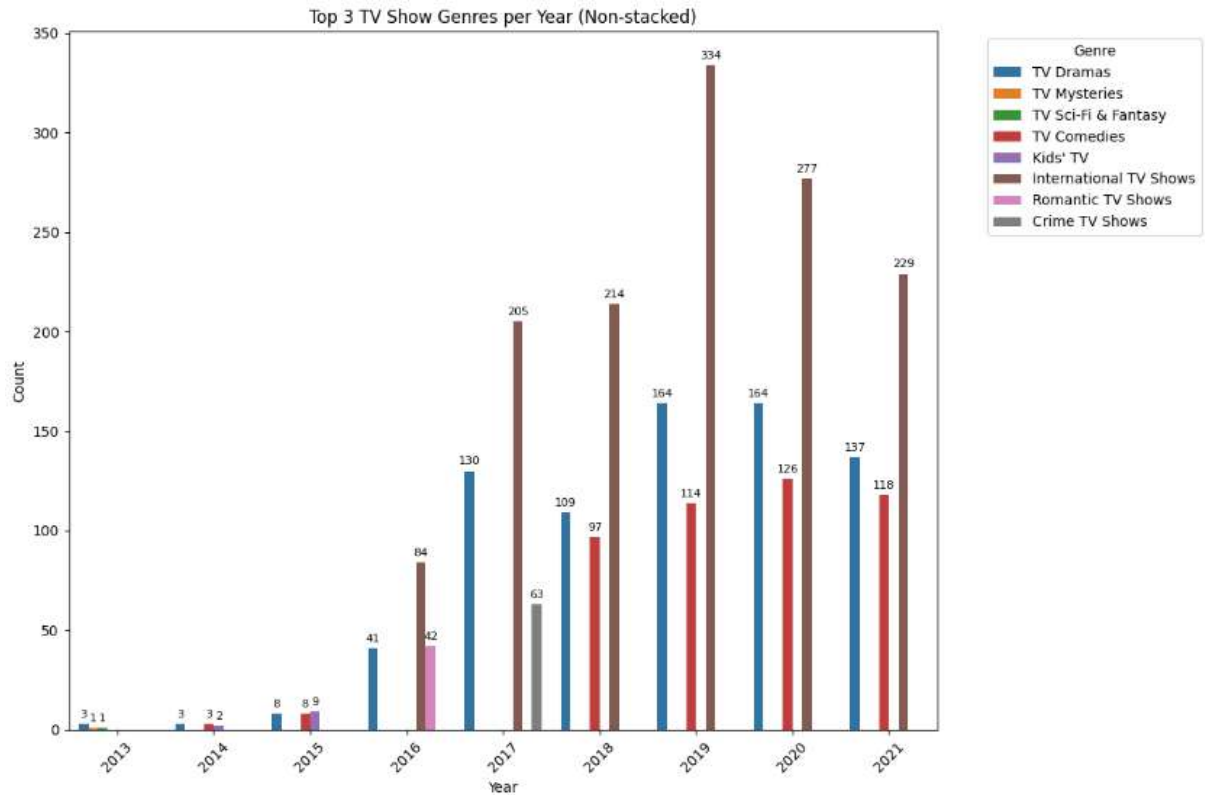
Genre seperti Children & Family Movies, Romantic Movies, dan Thrillers juga masuk dalam sepuluh besar, menandakan perhatian terhadap segmen keluarga, penikmat romansa, dan penggemar cerita menegangkan. Terakhir, genre Music & Musicals memiliki 375 judul.



Gambar 4.47 Visualisasi Top 3 Genres per Year

Visualisasi di atas menunjukkan bahwa sejak tahun 2018 hingga 2021, genre *International Movies* dan *Dramas* mendominasi platform Netflix. Pada tahun 2018, genre *International Movies* mencapai 668 judul, disusul oleh *Dramas* sebanyak 531 judul, dan *Comedies* dengan 297 judul. Popularitas genre *Dramas* dan *International Movies* tampak konsisten hingga tahun 2021, di mana keduanya masih berada di posisi teratas, meskipun terjadi sedikit penurunan jumlah judul dibanding tahun-tahun sebelumnya. Penurunan ini bisa dikaitkan dengan berbagai faktor, termasuk dampak pandemi global terhadap produksi film.

Genre lain yang secara konsisten masuk dalam tiga besar adalah *Comedies*, terutama dari tahun 2018 hingga 2020. Selain itu, *Documentaries* sempat mengalami peningkatan signifikan pada tahun 2017 (206 judul), namun tidak bertahan dalam posisi tiga besar di tahun-tahun berikutnya.

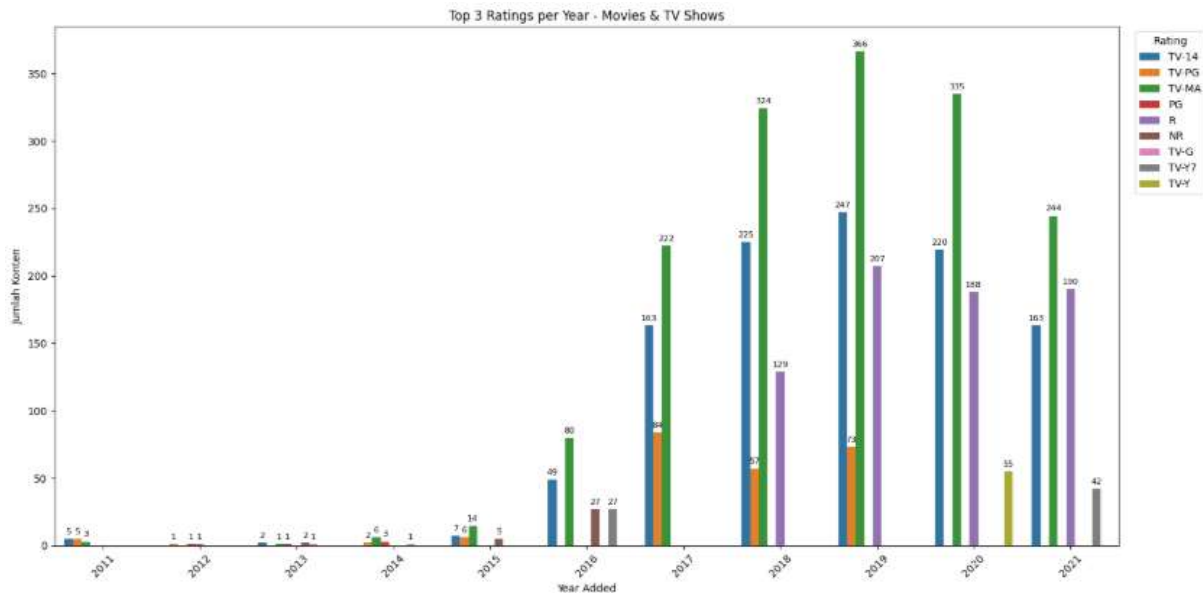


Gambar 4.48 Visualisasi Top 3 TV Show Genres per Year

Genre *International TV Shows* secara konsisten mendominasi sejak tahun 2016 dan mencapai puncaknya pada tahun 2019 dengan jumlah 334 tayangan, sebelum mengalami penurunan pada tahun 2020 dan 2021. Hal ini menegaskan peran penting konten internasional dalam strategi ekspansi global Netflix. Genre *TV Dramas* mulai menunjukkan peningkatan signifikan sejak tahun 2016, dan terus konsisten sebagai salah satu genre favorit dengan jumlah tayangan tertinggi kedua pada tahun 2019 dan 2020, masing-masing sebanyak 164 dan 126 tayangan.

Sementara itu, genre *TV Mysteries* mulai naik daun pada tahun 2017 dan berhasil mempertahankan eksistensinya dalam tiga besar hingga tahun 2021, dengan puncak tertinggi pada tahun 2020 sebanyak 137 tayangan. Genre ini menggambarkan ketertarikan audiens terhadap cerita yang penuh ketegangan dan teka-teki. Beberapa genre lain yang juga muncul namun tidak secara konsisten berada di posisi tiga besar adalah *Romantic TV Shows*, *Crime TV Shows*, dan *Kids' TV*

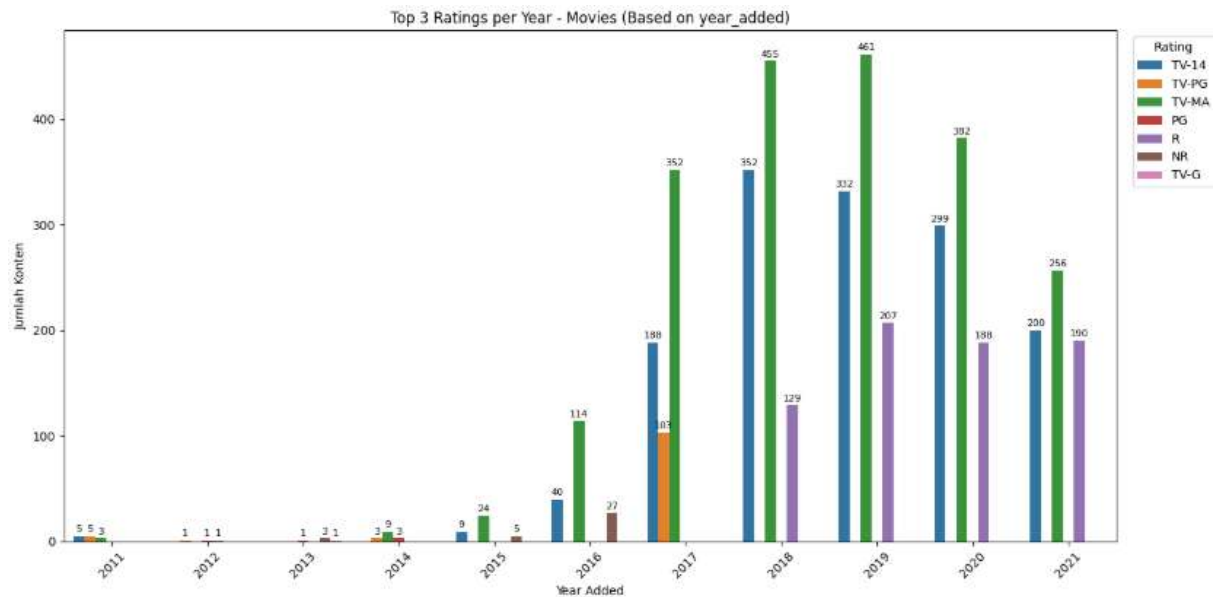
Gambar 4.49 menampilkan distribusi tiga peringkat usia (rating) teratas dari film dan acara TV yang ditambahkan ke Netflix setiap tahun dari 2011 hingga 2021. Visualisasi ini memberikan gambaran mengenai klasifikasi usia yang paling dominan dalam konten Netflix selama satu dekade terakhir.



Gambar 4.49 Visualisasi Top 3 Age Ratings per Year

Berdasarkan grafik di atas, terlihat bahwa rating TV-MA (Mature Audience Only) secara konsisten menempati posisi teratas dari tahun 2017 hingga 2021. Puncaknya terjadi pada tahun 2019 dengan jumlah 316 konten, menunjukkan bahwa Netflix banyak memproduksi atau mengakuisisi konten untuk audiens dewasa selama periode tersebut. Rating TV-14 juga mendominasi sejak tahun 2016 dan tetap tinggi hingga 2020, menandakan besarnya jumlah konten yang ditujukan untuk remaja ke atas. Tahun 2018 dan 2019 menjadi masa puncak bagi rating ini, masing-masing dengan jumlah 247 dan 207 konten.

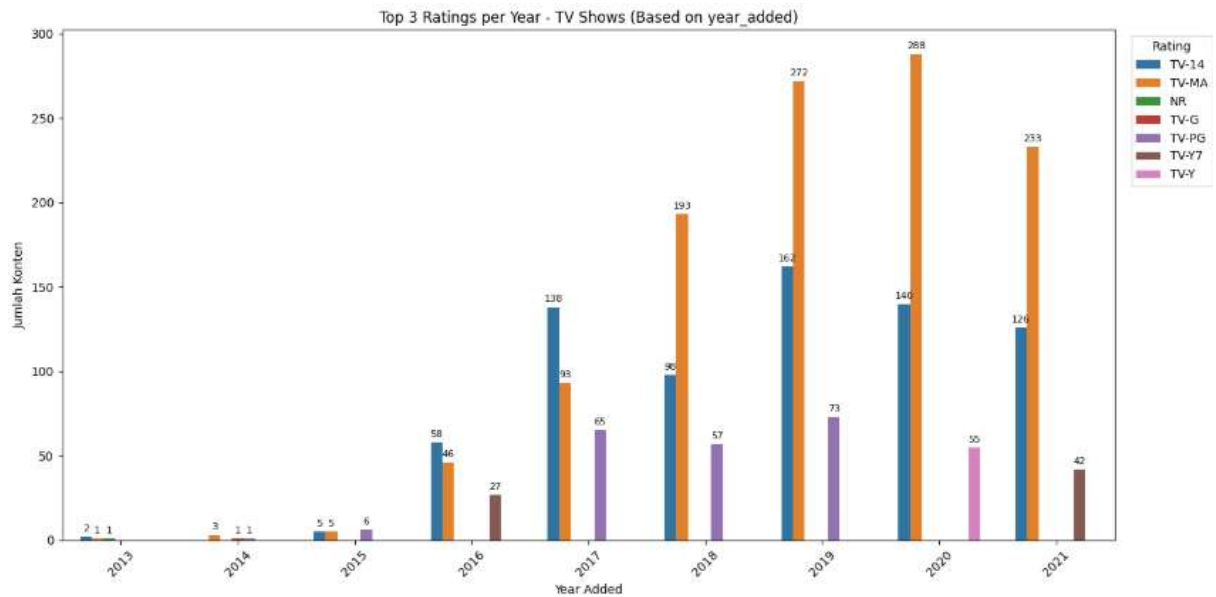
Sementara itu, rating TV-PG (Parental Guidance) juga sering masuk dalam tiga besar, terutama pada tahun 2019 dan 2020. Hal ini menunjukkan bahwa Netflix tetap mempertahankan konten ramah keluarga dan remaja meskipun tren konten dewasa meningkat. Rating lain seperti PG, R, dan TV-Y7 juga muncul di beberapa tahun, namun jumlahnya tidak sebesar tiga besar yang disebutkan. Hal ini mencerminkan adanya variasi konten untuk segmen penonton anak-anak, namun dengan jumlah yang relatif kecil.



Gambar 4.50 Visualisasi Top 3 Age Ratings Movies per Year

Visualisasi pada Gambar 4.50 menampilkan tiga klasifikasi usia teratas (age rating) pada konten film yang ditambahkan ke platform Netflix setiap tahun. Grafik ini menunjukkan bahwa rating TV-MA (konten untuk dewasa) secara konsisten mendominasi dari tahun ke tahun, diikuti oleh TV-14 dan TV-G. Hal ini mencerminkan strategi Netflix dalam menyediakan beragam konten yang menargetkan penonton remaja hingga dewasa.

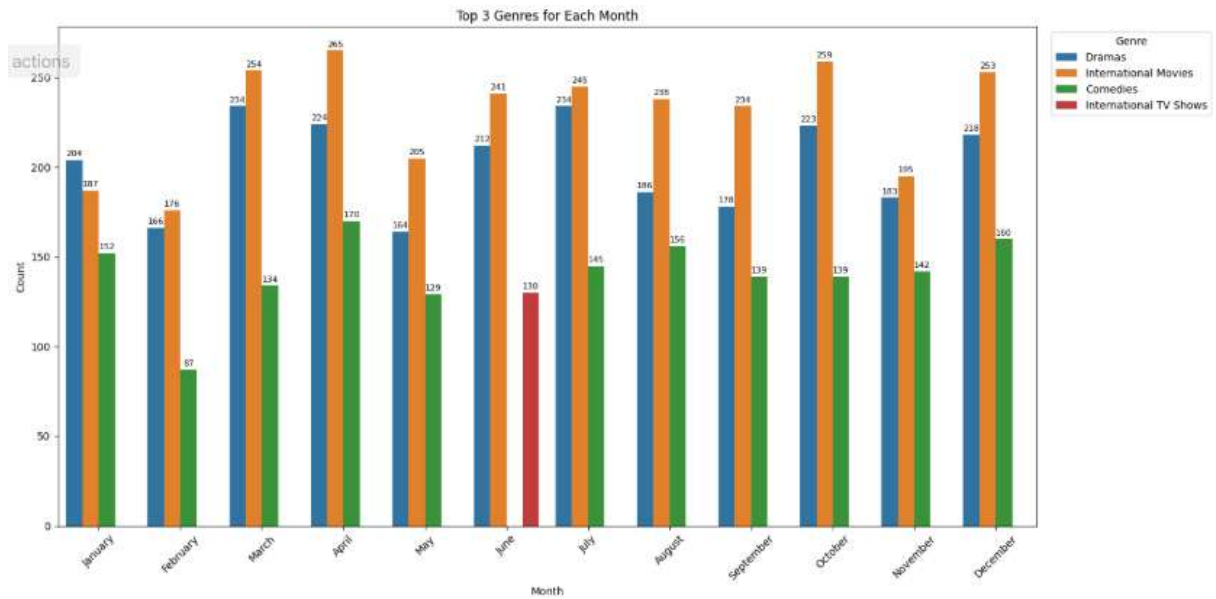
Puncak penambahan film dengan rating TV-MA terjadi pada tahun 2019 dan 2020, yang sejalan dengan tren meningkatnya konsumsi konten dewasa di platform tersebut. Sementara itu, rating TV-14 tetap konsisten tinggi, menunjukkan banyaknya film yang juga menyasar penonton remaja. Kehadiran rating TV-G (General Audience) mengindikasikan bahwa Netflix juga tetap mempertahankan sejumlah konten yang dapat dinikmati bersama keluarga.



Gambar 4.51 Visualisasi Top 3 Age Ratings TV Shows per Year

Dari grafik tersebut, dapat terlihat bahwa TV-MA (Mature Audience Only) secara konsisten mendominasi sejak tahun 2016 hingga 2021. Puncak penambahan konten dengan rating TV-MA terjadi pada tahun 2020 dengan 288 konten, disusul oleh tahun 2019 dengan 272 konten. Hal ini mengindikasikan bahwa Netflix banyak menambahkan konten yang ditujukan untuk penonton dewasa dalam beberapa tahun terakhir. Sementara itu, TV-14 juga menjadi salah satu kategori yang paling banyak muncul, menunjukkan banyaknya konten yang ditujukan untuk remaja ke atas. Penambahan konten dengan rating ini cenderung stabil, dengan angka tertinggi pada tahun 2019 sebanyak 162 konten.

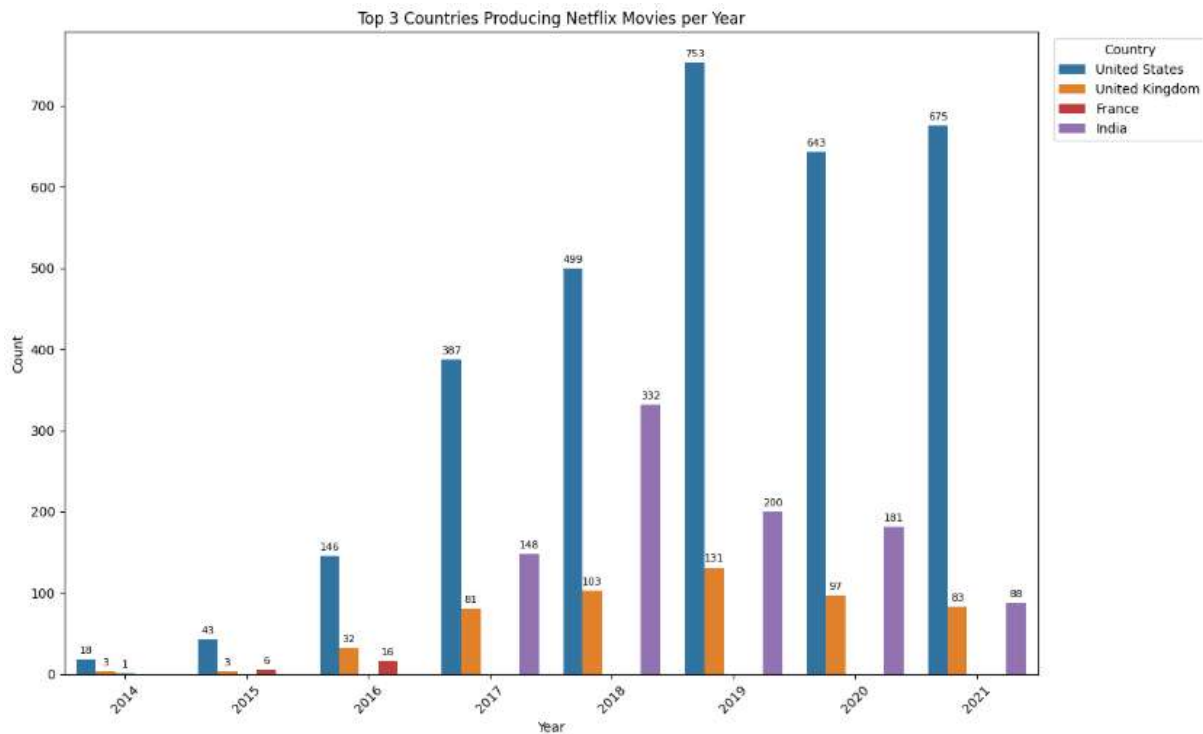
Selain kedua rating tersebut, TV-PG dan TV-Y7 juga muncul sebagai bagian dari tiga besar di beberapa tahun tertentu, khususnya pada tahun-tahun dengan peningkatan jumlah konten keluarga dan anak-anak. Misalnya, TV-Y7 memiliki peningkatan signifikan pada tahun 2020 dengan 55 konten, kemungkinan besar berkaitan dengan strategi Netflix dalam menghadirkan konten ramah keluarga selama masa pandemi.



Gambar 4.52 Visualisasi Top 3 Genres per Month

Berdasarkan grafik tersebut, genre *International Movies* tampak secara konsisten mendominasi di hampir seluruh bulan, dengan puncaknya pada bulan April dan Oktober. Dominasi genre ini menunjukkan fokus Netflix yang kuat dalam menghadirkan film-film dari berbagai negara sebagai bagian dari strategi globalisasi konten. Selain itu, genre *Dramas* juga menjadi salah satu yang paling sering muncul, mencerminkan minat pengguna yang tinggi terhadap narasi emosional dan mendalam. Dalam hampir setiap bulan, *Dramas* berada di urutan kedua setelah *International Movies*, menegaskan bahwa genre ini tetap menjadi salah satu daya tarik utama di platform Netflix.

Sementara itu, genre *Comedies* juga menunjukkan stabilitas yang cukup kuat dalam perilisannya, menempati posisi ketiga di sebagian besar bulan. Hal ini menandakan adanya permintaan yang konsisten terhadap konten hiburan ringan. Menariknya, genre *International TV Shows* hanya muncul sebagai tiga besar pada bulan Juni, yang mengindikasikan bahwa perilis serial internasional cenderung bersifat musiman atau disesuaikan dengan waktu tertentu. Secara keseluruhan, visualisasi ini menggambarkan bagaimana Netflix mengatur perilis genre secara strategis sepanjang tahun untuk menjaga keberagaman konten dan menjangkau preferensi pengguna dari berbagai segmen.

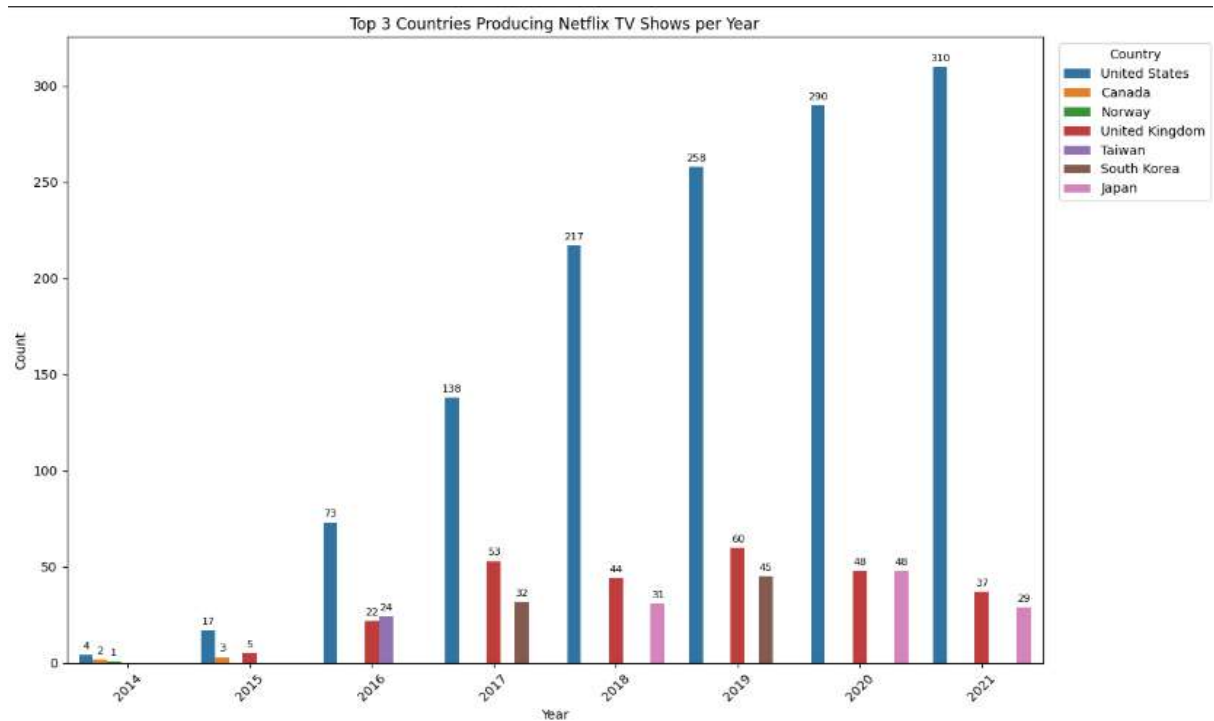


Gambar 4.53 Visualisasi Top 3 Movies Countries per Year

Amerika Serikat secara konsisten mendominasi sebagai negara dengan jumlah produksi film terbanyak pada platform Netflix. Dominasi ini semakin mencolok sejak tahun 2017, dengan puncaknya terjadi pada tahun 2018, di mana jumlah produksi film mencapai 753 judul. Tren ini menggambarkan kuatnya industri hiburan Amerika Serikat serta peran strategis negara tersebut dalam suplai konten global Netflix. Selain Amerika Serikat, negara India mulai menunjukkan kontribusi signifikan sejak tahun 2017, bahkan sempat menduduki posisi kedua pada tahun 2018 dengan 332 judul.

Hal ini menunjukkan pertumbuhan pesat industri perfilman India di platform global dan strategi Netflix dalam menyasar pasar Asia yang besar. Di sisi lain, Inggris dan Prancis tampil sebagai dua negara Eropa yang konsisten berada dalam tiga besar, dengan jumlah produksi yang stabil setiap tahunnya. Keberadaan negara-negara ini dalam grafik memperlihatkan pola globalisasi produksi konten yang dijalankan oleh Netflix, serta bagaimana platform ini secara aktif menggandeng industri perfilman dari berbagai belahan dunia untuk memperluas cakupan dan menjangkau audiens internasional. Visualisasi ini mendukung temuan sebelumnya bahwa

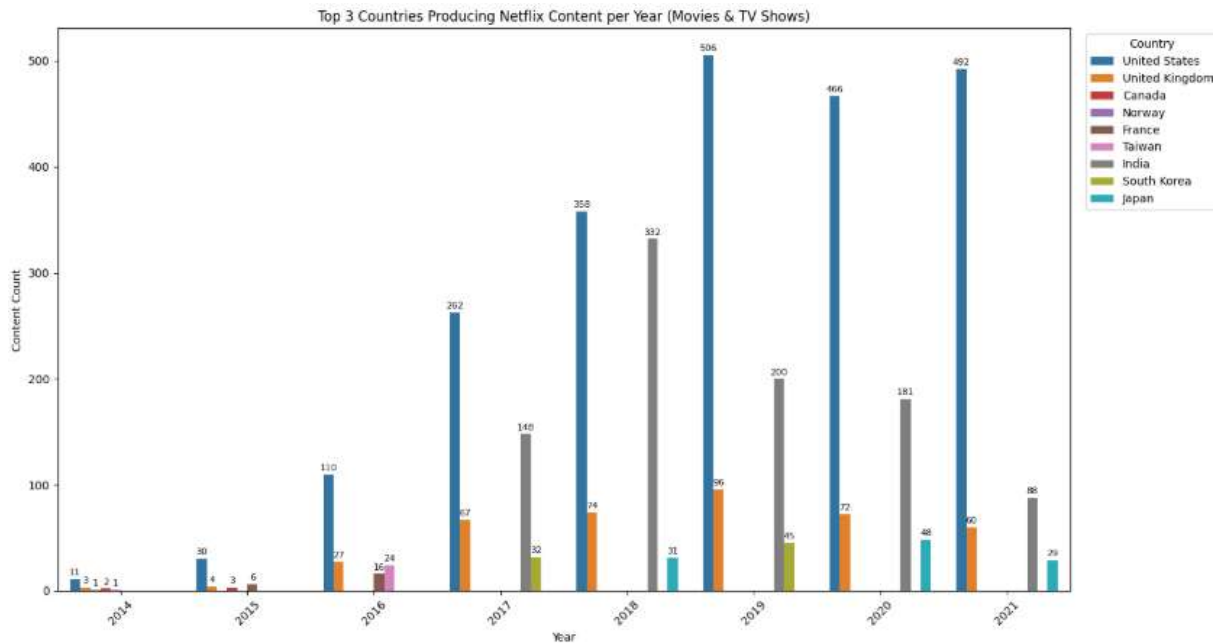
konten dari negara-negara seperti Amerika Serikat, India, dan Inggris merupakan tulang punggung dalam katalog film Netflix.



Gambar 4.54 Visualisasi Top 3 TV Shows Countries per Year

Gambar 4.54 memperlihatkan distribusi tiga negara teratas yang paling banyak memproduksi acara TV yang tersedia di platform Netflix setiap tahunnya hingga 2021. Terlihat bahwa Amerika Serikat secara konsisten menjadi negara dengan jumlah produksi acara TV terbanyak, bahkan sejak awal kemunculan Netflix sebagai platform streaming global. Angka produksi terus meningkat setiap tahun, dengan puncaknya pada tahun 2021 yang mencapai 310 judul. Fakta ini menunjukkan bahwa dominasi Amerika Serikat dalam konten netflix tidak hanya terbatas pada film, tetapi juga sangat kuat di ranah acara TV.

Selain Amerika Serikat, beberapa negara lain juga mencatat kontribusi signifikan. Inggris dan Kanada tampil sebagai dua negara yang paling konsisten berada di posisi tiga besar selama periode pengamatan, memperlihatkan kekuatan industri hiburan mereka di pasar internasional. Menariknya, beberapa negara Asia seperti Jepang, Korea Selatan, dan Taiwan mulai muncul secara konsisten sejak 2017, menandakan bahwa Netflix semakin mengakomodasi dan mempromosikan konten dari kawasan Asia. Hal ini tidak lepas dari tren globalisasi budaya pop Asia, seperti K-drama dan anime, yang kian diminati oleh audiens global.



Gambar 4.55 Visualisasi Top 3 Countries per Year

Gambar 4.55 menunjukkan tren produksi konten film dan acara TV Netflix dari tahun 2014 hingga 2021 berdasarkan negara. Amerika Serikat secara konsisten menjadi negara dengan jumlah konten terbanyak, mengalami lonjakan signifikan sejak 2016 dan mencapai puncaknya pada tahun 2019 dengan 506 konten. Selain AS, Inggris dan India juga menunjukkan kontribusi yang kuat. India mengalami peningkatan tajam pada tahun 2018 dengan 332 konten, mencerminkan pertumbuhan pasar Asia sebagai fokus ekspansi Netflix.

Tren ini menunjukkan bahwa Netflix semakin memperluas jangkauannya secara global. Negara-negara seperti Jepang, Korea Selatan, dan Kanada juga mulai menunjukkan pertumbuhan kontribusi, terutama setelah tahun 2017. Hal ini mencerminkan strategi Netflix untuk menghadirkan konten dengan preferensi wilayah masing-masing, sekaligus memperkaya katalog global mereka dengan keragaman budaya dan bahasa.

B. Analisis K-means dengan *elbow method* dan *silhouette score*

Setelah melakukan visualisasi umum, langkah selanjutnya adalah melakukan pengelompokan data untuk mengetahui segmentasi konten film berdasarkan tahun rilis dan durasi, maka dari itu akan dilakukan proses klusterisasi menggunakan algoritma K-Means. Sebelumnya, ditentukan jumlah kluster optimal menggunakan dua pendekatan, yaitu Elbow Method dan Silhouette Score. Disini telah dipilih dua variabel yaitu *release_year* dan *duration*.

Pemilihan kedua variabel ini bukan tanpa alasan, melainkan bertujuan untuk mengidentifikasi tren dan pola distribusi konten yang berkembang dari waktu ke waktu.

Fitur `release_year` digunakan untuk menangkap dinamika industri hiburan lintas tahun, memungkinkan kita untuk mengamati bagaimana karakteristik produksi konten berubah seiring perkembangan zaman. Misalnya, dalam beberapa tahun terakhir, terdapat kecenderungan bahwa durasi tayangan menjadi lebih singkat, diduga sebagai respons terhadap kebiasaan menonton masyarakat modern yang cenderung memilih konten instan dan cepat.

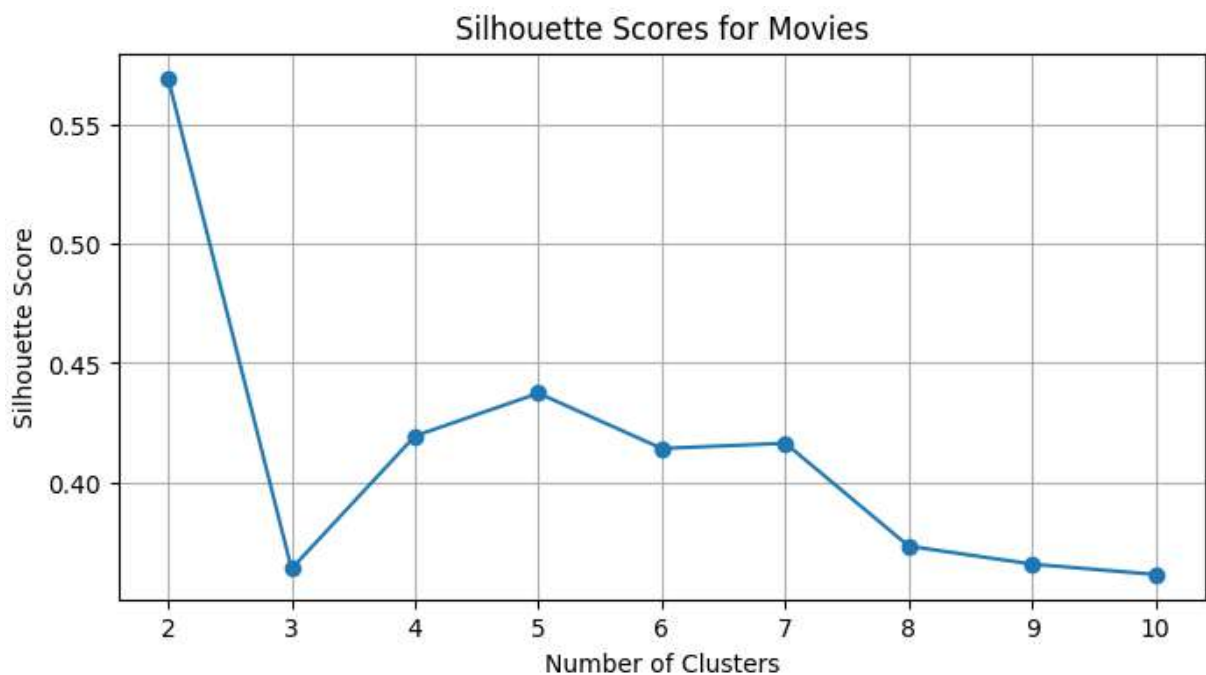
Sementara itu, fitur `duration` mewakili panjang durasi film atau episode yang dapat mencerminkan gaya penyutradaraan atau ekspektasi audiens pada periode tertentu. Kombinasi kedua fitur ini memungkinkan proses klusterisasi untuk tidak hanya memisahkan konten berdasarkan durasi atau tahun rilis secara individual, tetapi juga mengidentifikasi kelompok-kelompok yang mewakili era tertentu dalam industri hiburan. Sebagai contoh, film dari era 1990-an cenderung memiliki durasi yang berbeda dibandingkan dengan film yang diproduksi setelah tahun 2000-an, sehingga keduanya berpotensi tergolong ke dalam klaster yang berbeda.



Gambar 4.56 Visualisasi *elbow method* movies

Gambar 4.56 menunjukkan hasil plot *Elbow Method* pada data film. Dari grafik tersebut, terlihat bahwa penurunan nilai WCSS (*Within-Cluster Sum of Squares*) cukup tajam dari

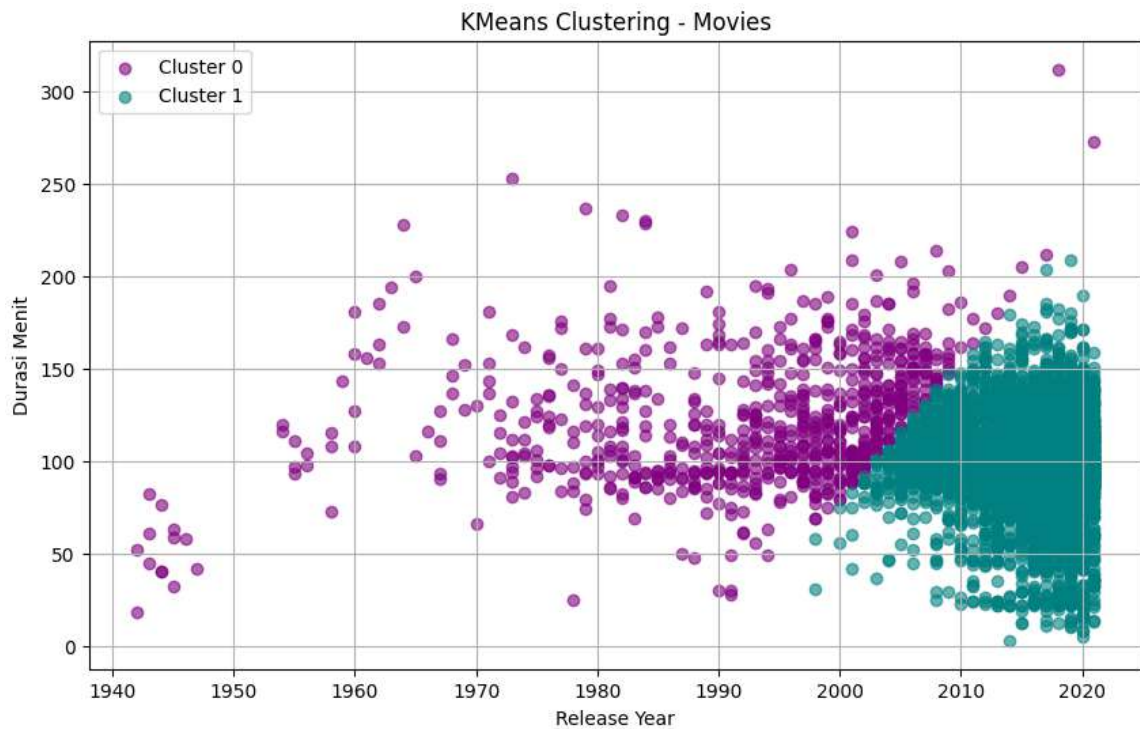
jumlah kluster 2 hingga 5. Setelah titik tersebut, penurunan menjadi lebih landai, yang menandakan bahwa menambah jumlah kluster selanjutnya tidak memberikan pengurangan WCSS yang signifikan. Dengan demikian, titik siku berada di sekitar kluster ke-4 atau ke-5, namun secara umum sulit untuk menetapkan satu titik pasti hanya dari grafik ini karena tidak terdapat sudut siku yang sangat jelas.



Gambar 4.57 Visualisasi *silhouette* analisis movies

Gambar 4.57 selanjutnya memperlihatkan nilai Silhouette Score dari jumlah kluster 2 hingga 10. Dari hasil ini, tampak bahwa nilai tertinggi berada pada jumlah kluster 2, dengan skor 0.57, yang menunjukkan bahwa pembagian dua kluster memberikan pemisahan data paling optimal berdasarkan jarak antar kluster dan kepadatan internal. Setelah nilai 2, skor mengalami penurunan dan berfluktuasi, sehingga semakin banyak kluster tidak menjamin kualitas segmentasi yang lebih baik.

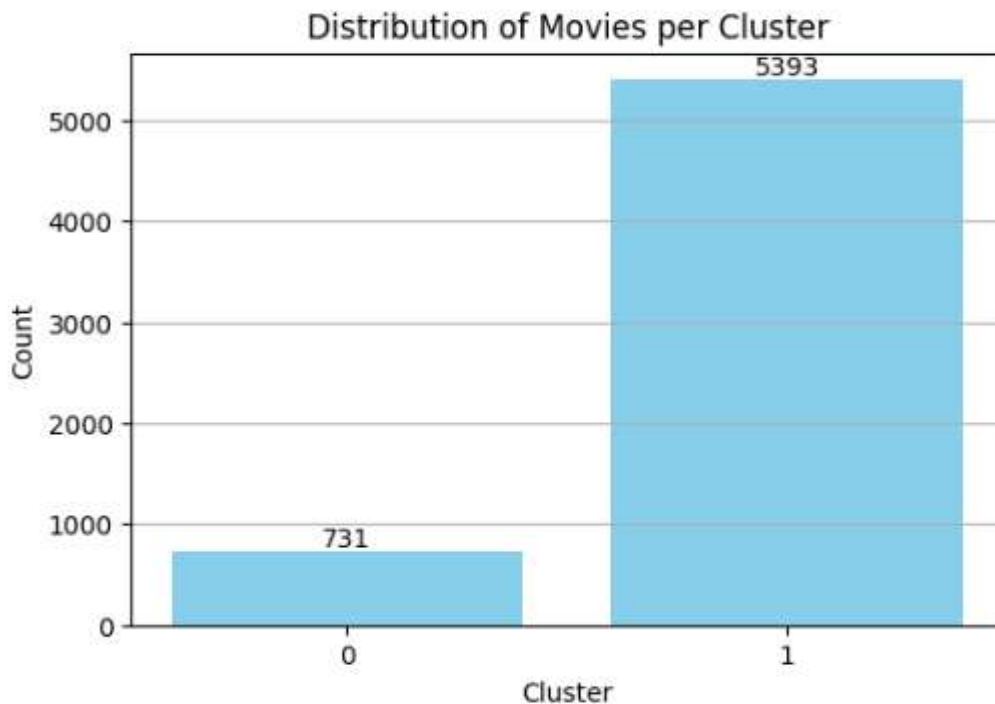
Karena data yang digunakan dalam klustering telah dipisahkan berdasarkan tipe (*Movie*), maka jumlah data yang digunakan menjadi lebih kecil dibandingkan keseluruhan dataset. Berdasarkan hal ini, pendekatan Silhouette Score lebih diutamakan untuk menentukan jumlah kluster optimal karena lebih sesuai pada kasus dengan data terbatas dan dua fitur numerik, sebagaimana disarankan oleh literatur (Kassambara, 2017).



Gambar 4.58 Visualisasi K-means Movies

Setelah ditentukan bahwa jumlah kluster optimal adalah 2, maka proses K-Means dijalankan dengan nilai $k = 2$. Hasil visualisasi kluster ditampilkan pada Gambar 4.X. Titik-titik mewakili film yang dipetakan berdasarkan tahun rilis pada sumbu X dan durasi pada sumbu Y. Dari grafik tersebut, terlihat bahwa kedua kluster membentuk pola yang cukup terpisah: Cluster 0 mencakup film dengan tahun rilis yang lebih lama dengan variasi durasi yang lebih luas, sedangkan Cluster 1 mendominasi film-film modern dengan durasi yang relatif lebih konsisten di sekitar 90–120 menit. Hal ini mencerminkan tren konten film yang semakin terstandarisasi dari segi durasi seiring dengan perkembangan waktu.

Langkah lanjutan setelah melakukan proses klusterisasi pada data film menggunakan algoritma K-Means dan menetapkan jumlah kluster optimal sebanyak 2 berdasarkan nilai Silhouette Score tertinggi, dilakukan visualisasi distribusi jumlah film dalam masing-masing kluster. Hasil distribusi tersebut ditunjukkan pada Gambar 4.59.

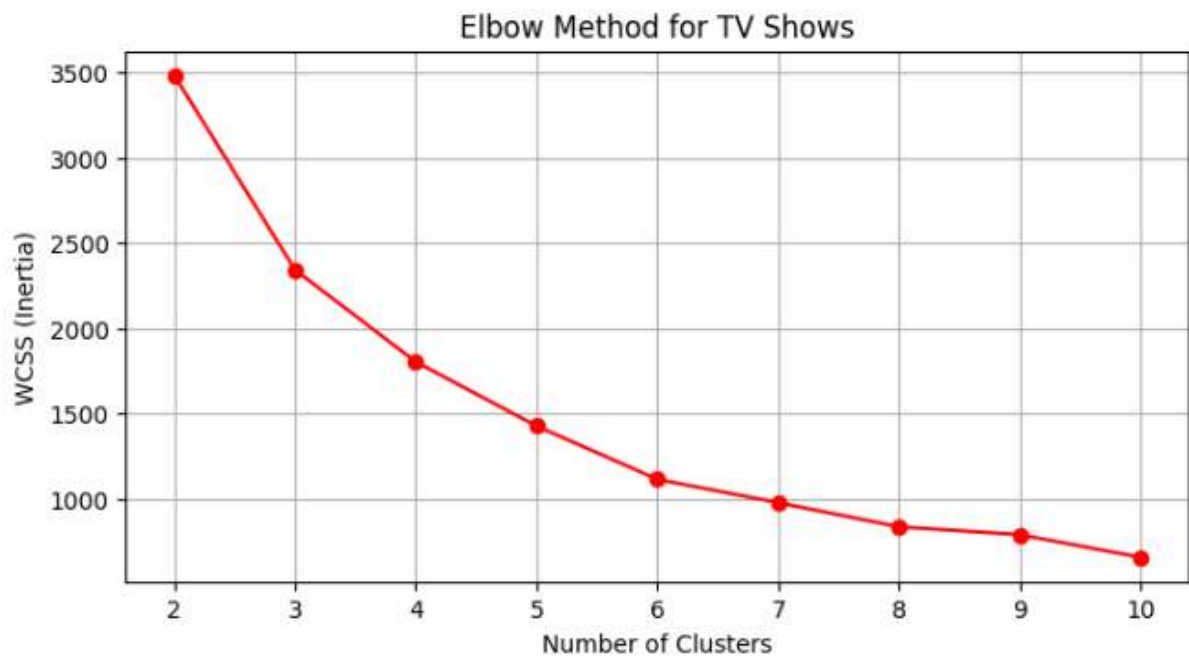


Gambar 4.59 Distribusi Film per Klaster Movies

Dari visualisasi pada Gambar 4.59, terlihat bahwa mayoritas film tergolong ke dalam Cluster 1 dengan total 5.393 film, sedangkan Cluster 0 hanya mencakup 731 film. Hal ini menunjukkan adanya perbedaan yang cukup mencolok dalam karakteristik antara kedua klaster. Cluster 1 kemungkinan besar merepresentasikan film-film modern dengan durasi yang relatif konsisten, sementara Cluster 0 mencakup film-film lama dengan rentang durasi yang lebih bervariasi. Perbedaan ini mengindikasikan segmentasi yang kuat dalam pola distribusi konten film berdasarkan tahun rilis dan durasi.

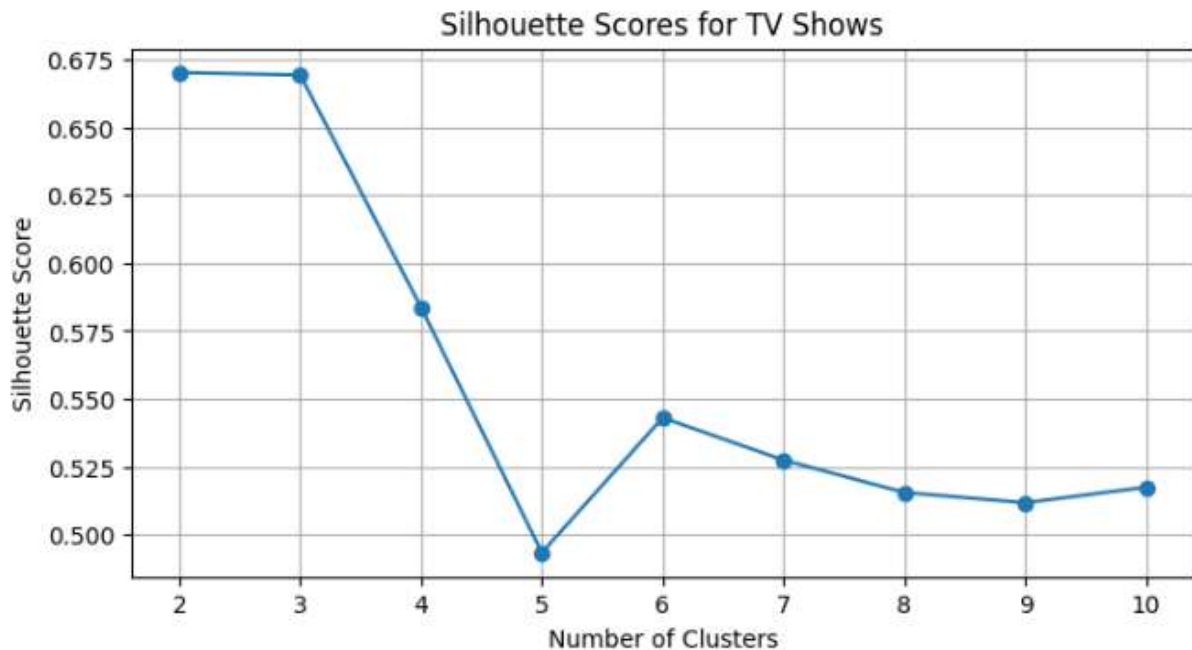
Pendekatan yang sama juga diterapkan pada subset data bertipe TV Show. Tujuannya tetap sama, yaitu untuk mengidentifikasi segmentasi konten berdasarkan tahun rilis dan durasi episode, guna melihat pola-pola yang mungkin muncul dalam produksi TV Show di Netflix.

Proses penentuan jumlah klaster optimal kembali menggunakan dua pendekatan, yaitu Elbow Method dan Silhouette Score. Namun, seperti pada data *Movie*, pendekatan Elbow Method juga mengalami kendala dalam hal kejelasan titik siku karena keterbatasan variasi dan jumlah data.



Gambar 4.60 Visualisasi *elbow method* TV Shows

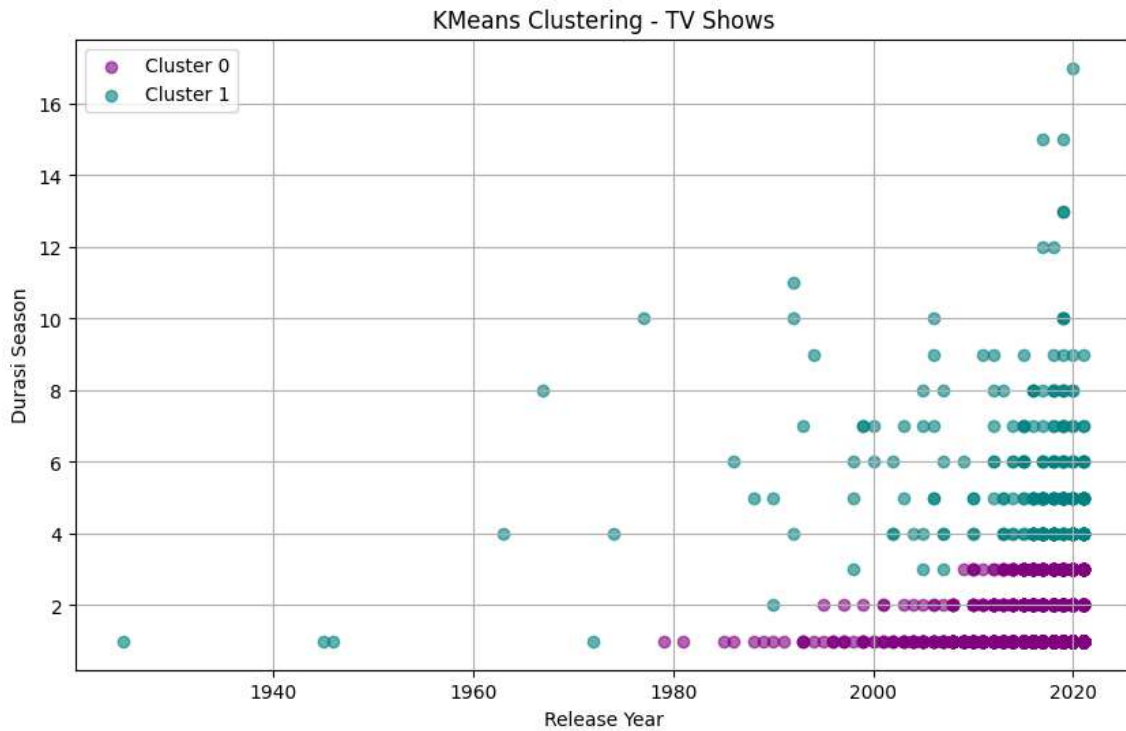
Sebagaimana terlihat pada Gambar 4.60, grafik Elbow menunjukkan penurunan nilai WCSS yang signifikan antara kluster 2 hingga 5, kemudian cenderung melandai. Meskipun pola ini menyerupai yang terjadi pada data *Movie*, namun tidak terdapat titik siku yang benar-benar jelas. Hal ini kembali memperkuat alasan bahwa Elbow Method kurang ideal jika digunakan secara tunggal dalam kondisi dataset yang terbatas.



Gambar 4.61 Visualisasi *silhouette* analisis TV Shows

Sebaliknya, Gambar 4.61 yang memperlihatkan hasil Silhouette Score menunjukkan bahwa nilai tertinggi berada pada jumlah kluster 2, dengan skor mendekati 0.675. Nilai ini bahkan lebih tinggi dibandingkan hasil Silhouette pada *Movie*, yang mengindikasikan bahwa pembagian dua kluster untuk data TV Show memberikan segmentasi yang sangat baik, dengan pemisahan kluster yang cukup tegas dan kohesi internal yang tinggi.

Dengan mempertimbangkan kembali ukuran data dan kompleksitas fitur yang terbatas, pendekatan Silhouette Score tetap menjadi acuan utama dalam pemilihan jumlah kluster untuk data (Kassambara, 2017).

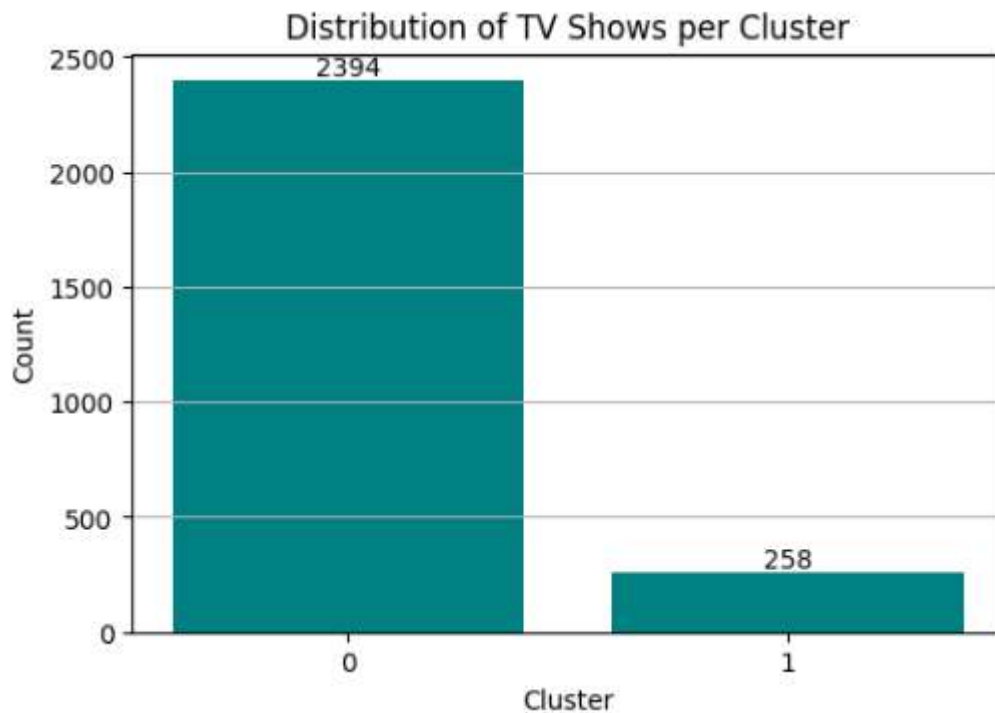


Gambar 4.62 Visualisasi K-means TV Shows

Setelah ditetapkan jumlah kluster optimal adalah 2, proses K-Means dijalankan untuk data TV Show. Hasilnya divisualisasikan pada Gambar 4.62, dengan sumbu X mewakili tahun rilis dan sumbu Y mewakili durasi episode (jumlah episode dari sebuah acara TV). Hasil visualisasi menunjukkan pola segmentasi yang cukup menarik:

Cluster 0 didominasi oleh TV Show modern (rilis setelah tahun 2000) dengan durasi sangat pendek, seringkali hanya terdiri dari 1 hingga 3 episode. Ini mengindikasikan adanya tren mini-series atau konten terbatas yang semakin populer dalam beberapa tahun terakhir. Cluster 1, sebaliknya, mencakup TV Show dengan durasi lebih panjang dan beragam, serta waktu rilis yang lebih menyebar sejak dekade 1980-an hingga saat ini. Kluster ini cenderung mewakili konten dengan jumlah episode yang lebih bervariasi.

Dengan demikian, segmentasi yang dihasilkan dari klasterisasi ini memperlihatkan adanya perbedaan pendekatan produksi TV Show yang dapat menjadi cerminan strategi konten Netflix dalam memenuhi kebutuhan pasar global yang semakin dinamis.



Gambar 4.63 Pembagian visualisasi sesuai kluster TV Shows

Hasil visualisasi menunjukkan bahwa sebagian besar TV Show tergabung dalam Cluster 0 dengan jumlah 2.394 judul, sedangkan Cluster 1 hanya mencakup 258 judul. Perbedaan yang signifikan ini menandakan bahwa mayoritas TV Show memiliki karakteristik yang mirip, baik dari segi tahun rilis maupun jumlah durasi rata-rata per episode. Hal ini sejalan dengan kecenderungan konten TV Show modern yang lebih seragam dan terstandardisasi, seperti penggunaan format *mini-series* atau *limited series* yang memiliki jumlah episode dan durasi yang relatif konsisten..

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini menganalisis 8.807 data tayangan dari platform Netflix, yang terdiri dari 6.126 film (69,7%) dan 2.664 acara TV (30,3%). Tujuan dari penelitian ini adalah untuk mengidentifikasi pola dan tren distribusi konten Netflix berdasarkan eksplorasi variabel-variabel yang tersedia dalam dataset. Analisis dilakukan dengan pendekatan Exploratory Data Analysis (EDA), visualisasi data, serta penerapan algoritma K-Means Clustering guna memperoleh segmentasi konten secara menyeluruh.

Hasil penelitian menunjukkan bahwa Amerika Serikat merupakan negara asal dengan jumlah konten terbanyak, disusul oleh India, Inggris, Kanada, dan Jepang. Genre “International Movies” dan “Dramas” merupakan yang paling dominan pada film, sementara “International TV Shows” dan “TV Dramas” mendominasi kategori TV Show. Untuk klasifikasi usia, konten dengan rating TV-MA dan TV-14 paling sering muncul, menandakan bahwa mayoritas tayangan ditujukan untuk penonton dewasa dan remaja.

Dari sisi waktu perilisan, film paling banyak dirilis pada bulan Juli, bertepatan dengan musim panas (summer), sedangkan TV Show cenderung dirilis pada bulan Desember, yang berada pada musim dingin. Hal ini mengindikasikan adanya pola strategi perilisan konten yang disesuaikan dengan momentum musiman. Selain itu, terdapat tren peningkatan produksi konten internasional pada beberapa tahun terakhir, yang mendukung upaya ekspansi global Netflix.

Pada tahap klasterisasi, proses K-Means dilakukan secara terpisah untuk film dan TV Show. Untuk film, terbentuk dua klaster utama: klaster pertama berisi film-film lama dengan variasi durasi yang lebih luas, sedangkan klaster kedua didominasi oleh film-film modern dengan durasi yang lebih konsisten di sekitar 90–120 menit. Pada TV Show, klaster pertama berisi sebagian besar tayangan berdurasi pendek (1–3 episode), umumnya berupa mini-series modern, sedangkan klaster lainnya berisi konten dengan jumlah episode lebih banyak dan tahun rilis yang lebih bervariasi. Nilai silhouette score tertinggi pada clustering TV Show mencapai 0,67 yang menunjukkan bahwa segmentasi dua klaster memberikan hasil pemisahan yang cukup baik.

Secara keseluruhan, penelitian ini berhasil mengidentifikasi tren dan pola distribusi konten di platform Netflix dari berbagai dimensi, termasuk jenis tayangan, genre, waktu rilis, klasifikasi usia, dan segmentasi berdasarkan karakteristik produksi. Temuan ini diharapkan dapat memberikan kontribusi dalam pemahaman strategi konten berbasis data, serta mendukung penyusunan keputusan dalam distribusi, personalisasi, dan pengembangan konten digital pada industri hiburan global.

5.2 Saran

Harapannya, penelitian ini dapat menjadi kontribusi bagi komunitas ilmiah dan industri hiburan digital. Dengan menyajikan temuan yang mendalam dan analisis yang teliti, penelitian ini diharapkan menjadi landasan yang kokoh bagi studi lanjutan dalam bidang ini. Keberhasilan penelitian ini juga diharapkan menginspirasi peneliti lain untuk mengeksplorasi aspek-aspek baru dalam analisis industri hiburan digital.

Selain itu, diharapkan bahwa hasil penelitian ini dapat menjadi rujukan bagi platform streaming lainnya dalam merancang strategi konten yang lebih terarah. Dengan memahami preferensi pengguna berdasarkan asal negara dan jenis konten, platform dapat meningkatkan daya saing dan memperluas jangkauan pasar.

Untuk penelitian lanjutan, disarankan beberapa hal berikut:

1. Menggunakan dataset dengan periode waktu yang lebih panjang untuk mengidentifikasi tren jangka panjang.
2. Melakukan analisis pada platform streaming lain untuk perbandingan strategi konten.
3. Menggunakan algoritma analisis data yang lebih kompleks, seperti machine learning, untuk prediksi minat pengguna.
4. Merancang dashboard visualisasi yang lebih interaktif dan informatif untuk mempermudah interpretasi data.

Dengan menjadikan penelitian ini sebagai referensi yang andal, diharapkan peneliti, akademisi, dan praktisi di masa mendatang dapat mengembangkan studi yang lebih inovatif dan berkontribusi pada kemajuan industri hiburan digital secara global.

DAFTAR PUSTAKA

- Alaskar, T. H., Alsadi, A. K., Aloulou, W. J., & Ayadi, F. M. (2024). Big Data Analytics, Strategic Capabilities, and Innovation Performance: Mediation Approach of Organizational Ambidexterity. *Sustainability*, 16(12), 5111. <https://doi.org/10.3390/su16125111>
- Anggrahini, D., Kurniati, N., Prasanto, A. D., & Yaqin, M. A. (2020). Implementasi Data Analitik Untuk Penentuan Jadwal Pemeliharaan Mesin Pada Perusahaan Manufaktur Kontinu Skala Besar. *Sisfo*, 09(03). <https://doi.org/10.24089/j.sisfo.2020.05.003>
- Anggraini, D., Hapsari, P., Adhayanti, N., & Susiloatmaja, R. (2023). *Analisis PIECES untuk Evaluasi Layanan Aplikasi Disney+ Hotstar* (Vol. 4, Issue 3).
- Anisa, A. I. N. A., & Novita, N. (2023). Penggunaan Data Analytics Terhadap Peningkatan Deteksi Kecurangan Pada Pemeriksaan Eksternal. *Journal of Economic, Management, Accounting and Technology*, 6(1), 34–45. <https://doi.org/10.32500/jematech.v6i1.3435>
- Arora, S. K. (2020). *What Is Data Analysis? Methods, Technique, & Tools*. Hackr.Io.
- Celicia, S., #1, D., Bunyamin, H., & Budi, S. (2022). *Penerapan Data Science pada Analisis Data Acara TV dan Film pada Aplikasi Layanan Streaming* (Vol. 4). www.kaggle.com/shivamb/netflix-shows
- Dangeti, P. (2017). *Statistics for Machine Learning*. Packt Publishing Ltd.
- Deloitte. (2020). *How digital transformation can lead to sustainable high performance*.
- Feng, L., Sun, B., Wang, K., & Tsai, S.-B. (2018). An Empirical Study on the Design of Digital Content Products from a Big Data Perspective. *Sustainability*, 10(9), 3092. <https://doi.org/10.3390/su10093092>
- Fernandes Andry, J., Tannady, H., Ivana Limawal, I., Dwinoor Rembulan, G., & Farady Marta, R. (2021). BIG DATA ANALYSIS ON YOUTUBE WITH TABLEAU. *Journal of Theoretical and Applied Information Technology*, 99, 22. www.jatit.org
- Gudivada, V. N. (2017). Data Analytics. In *Data Analytics for Intelligent Transportation Systems* (pp. 31–67). Elsevier. <https://doi.org/10.1016/B978-0-12-809715-1.00002-X>
- Hermawan, F. (2022). Analysis of interest of the YouTube platform users as a medium of digital communication today. *Jurnal Manajemen*.
- Iskandar, B. J., & Sutanto, E. M. (2022). Analysis of movie genre availability, service quality, and movie quality on Netflix user satisfaction in Indonesia. *ITEJ (Information Technology Engineering Journals)*, 7(1).

- Jollyta, D., Efendi, S., Zarlis, M., & Mawengkang, H. (2019). Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1, 918. <https://doi.org/10.30645/senaris.v1i0.100>
- Kassambara, A. (2017). *Multivariate Analysis I Practical Guide To Cluster Analysis in R Unsupervised Machine Learning*. <http://www.sthda.com>
- Mulyandi, M. R., Angela, M., & Selyano, N. (2021). *Analysis of Disneyplus Hotstar Indonesia Youtube Advertising Version "Choose All, For All" During the Covid 19 Pandemic in July to December 2020*. <http://e-journal.stie-kusumanegara.ac.id>
- Mulyati, S., Noorhadi, D. F., Chaerulisma, H. F., & Setiani, N. (2025). Stunting Incidence Segmentation: A Cluster Analysis Approach and Targeted Intervention Strategies. *International Journal of Computational Methods and Experimental Measurements*, 13(1), 125–132. <https://doi.org/10.18280/ijcmem.130113>
- Nasir, N., & Sukmawati, S. (2023). Analysis of Research Data Quantitative and Qualitative. *Edumaspul: Jurnal Pendidikan*.
- Nurhayati, A. (2019). A Survey on Big Data in the Media and Entertainment Industry. *Information Technology Engineering Journals. ITEJ (Information Technology Engineering Journals)*, 7(1).
- Rachel, S. (2018). *Data Visualization*.
- Rick D. (2018). *Ways through which Technology is Making Businesses More Efficient*.
- Sahoo*, K., Samal, A. K., Pramanik, J., & Pani, S. K. (2019). Exploratory Data Analysis using Python. *International Journal of Innovative Technology and Exploring Engineering*, 8(12), 4727–4735. <https://doi.org/10.35940/ijitee.L3591.1081219>
- SGA Knowledge Team. (2023). *How is the Media and Entertainment Industry Leveraging Big Data Analytics for Disruptive Growth?* Sganalytics.
- Statista. (2022). *Netflix Sheds Subscribers for Second Quarter in a Row, Netflix's Subscriber Growth*.
- Sun, X. (2022). Two Decades and More: Evolution of Netflix's Business Models and Outlook of Future. *SHS Web of Conferences*, 148, 03032. <https://doi.org/10.1051/shsconf/202214803032>

LAMPIRAN