

Laporan UTS Praktikum Data Analysis



Nama Anggota :

41425064 Yosevyn Sipahutar

41425081 Lukas Simatupang

41425083 Inez Cecilia Tiurma

INSTITUT TEKNOLOGI DEL

FAKULTAS VOKASI

Sarjana Terapan Teknologi Rekayasa

Perangkat Lunak

BAB I

PENDAHULUAN

1. Latar Belakang

Industri streaming digital berkembang sangat cepat dalam 10 tahun terakhir, dan Netflix menjadi salah satu platform yang paling berpengaruh. Dataset publik dari Netflix berisi informasi lengkap tentang ribuan film dan serial, seperti negara asal, genre, tahun rilis, dan jenis tayangan. Analisis ini bertujuan untuk melihat bagaimana konten didistribusikan, tren jumlah produksi tiap tahun, serta ciri-ciri tayangan yang tersedia di Netflix. Adapun alasan dilakukan analisis ini:

- Industri streaming terus berkembang, dan jumlah konten Netflix semakin banyak.
- Penting untuk memahami pola produksi dan distribusi konten agar bisa mengikuti arah perkembangan industri.
- Dataset Netflix dipilih karena datanya lengkap dan sesuai dengan tren bisnis digital saat ini.

2. Tujuan

Tujuan dari analisis ini adalah untuk memperoleh pemahaman yang lebih mendalam mengenai karakteristik dan perkembangan konten di platform Netflix. Secara khusus, analisis ini bertujuan untuk:

1. **Mengidentifikasi tren jumlah tayangan** yang dirilis oleh Netflix dari tahun ke tahun, guna melihat pola pertumbuhan maupun penurunan produksi konten dalam kurun waktu tertentu.
2. **Mengetahui genre dan negara asal produksi yang paling dominan**, sehingga dapat menggambarkan kecenderungan serta distribusi konten secara global.
3. **Menganalisis durasi rata-rata antara film dan serial**, untuk memahami perbedaan karakteristik dari kedua jenis tayangan tersebut.
4. **Menyajikan temuan dan insight analisis dalam bentuk visualisasi interaktif** melalui dashboard Tableau, agar memudahkan interpretasi data serta mendukung pengambilan keputusan berbasis data secara efektif.

3. Ruang Lingkup (Scope)

Adapun ruanglingkup dari analisis ini adalah sebagai berikut:

- a) Analisis dilakukan pada dataset publik “Netflix Movies and TV Shows” dari platform Kaggle.
- b) Alat yang digunakan: File dataset dengan format .csv dan Tableau

BAB II

CRISP-DM

Cross-Industry Standard Process for Data Mining

A. Business Understanding

Tahap Business Understanding merupakan langkah awal yang sangat krusial dalam proses analisis data, karena menentukan arah, fokus, dan relevansi dari keseluruhan penelitian. Dalam proyek ini, analisis diarahkan untuk menjawab beberapa pertanyaan utama yang berkaitan dengan pemanfaatan data Netflix, yaitu:

1. Apa permasalahan atau pertanyaan utama yang ingin dijawab melalui data ini?
2. Apakah Netflix dapat dimanfaatkan sebagai media pembelajaran, jika ditinjau dari genre konten dan rating tayangannya?
3. Mengingat Netflix memiliki beragam jenis konten — seperti film, serial, dan dokumenter — dengan berbagai kategori, apakah terdapat proporsi konten yang bersifat edukatif? Jika ada, untuk kelompok usia mana konten tersebut paling sesuai?

Untuk menjawab pertanyaan-pertanyaan tersebut, digunakan pendekatan sistematis berdasarkan metodologi CRISP-DM, yang mencakup tahapan pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyajian hasil. Analisis yang dilakukan berfokus pada Analisis Deskriptif dan Visualisasi Data terhadap film dan serial di platform Netflix. Tujuannya adalah untuk menggali informasi terkait tren jumlah tayangan dari tahun ke tahun, distribusi genre dan negara asal, serta karakteristik konten berdasarkan durasi dan klasifikasi usia. Selain itu, analisis ini juga bertujuan untuk mengidentifikasi potensi konten edukatif yang tersedia di Netflix. Hasil dari analisis ini disajikan dalam bentuk visualisasi interaktif menggunakan Tableau, sehingga dapat memberikan insight yang lebih mudah dipahami.

B. Data Understanding

Deskripsi Dataset: Dataset berisi ribuan entri tayangan di Netflix dengan berbagai kolom informasi berikut:

Kolom	Tipe Data	Deskripsi	Kegunaan Analisis
Show ID	String	ID unik untuk setiap tayangan (misal: s1, s2)	Identifier utama (primary key)
Type	String	Jenis tayangan ('Movie' / 'TV Show')	Membandingkan proporsi film vs serial
Title	String	Judul tayangan	Identifikasi atau label visual

Kolom	Tipe Data	Deskripsi	Kegunaan Analisis
Director	String	Nama sutradara	Analisis produktivitas sutradara
Cast	String	Daftar aktor utama	Analisis kolaborasi atau popularitas aktor
Country	String	Negara produksi	Analisis distribusi geografis (map chart)
Date_added	Date	Tanggal tayangan ditambahkan ke Netflix	Analisis tren penambahan konten
Release_year	Number	Tahun rilis tayangan	Analisis tren produksi
Rating	String	Klasifikasi usia (PG-13, TV-MA, dll.)	Analisis segmen audiens
Duration	String	Lama tayangan ('90 min' / '1 Season')	Analisis rata-rata durasi
Listed_in	String	Genre atau kategori ('Drama, Romance')	Analisis distribusi genre
Description	String	Ringkasan sinopsis	Analisis teks opsional

C. Data Preparation

Tahapan Pembersihan dan Transformasi Data:

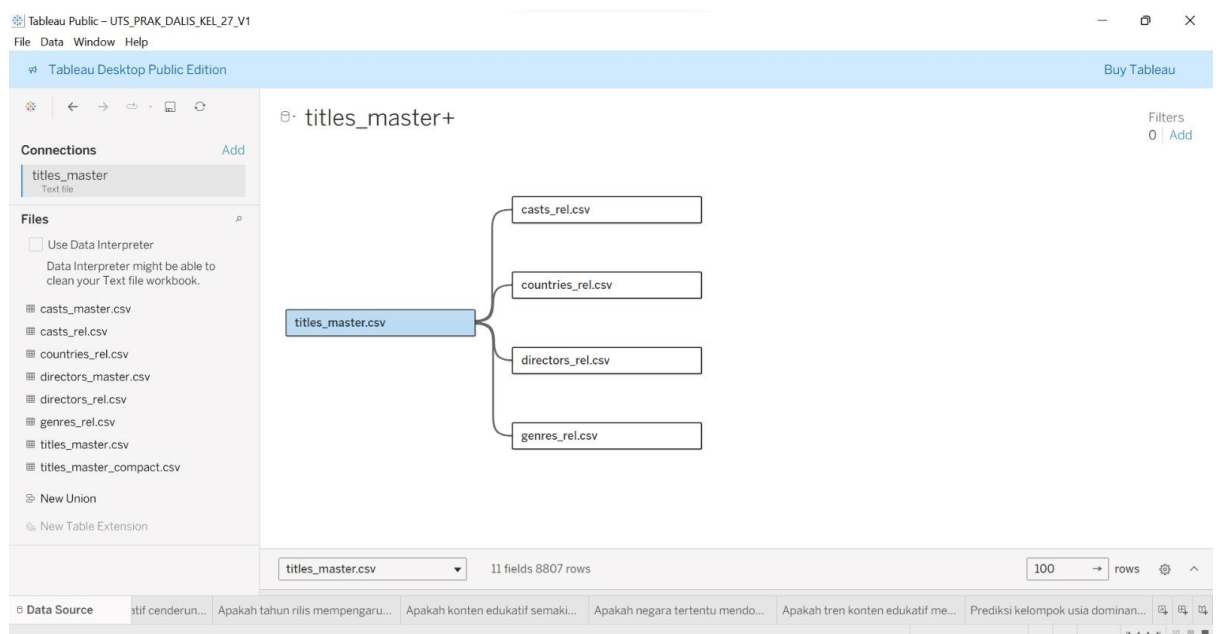
Tahap Cleaning Data:

1. **Kebocoran nilai pada kolom Rating dan Duration** — memperbaiki *miscollumn* untuk memastikan data berada pada kolom yang tepat.
2. **Kolom Country masih ada beberapa negara yang tergabung** — memisahkan data yang digabung agar setiap entri merepresentasikan satu nilai yang valid.
3. **Menangani kolom Director, Cast, dan Country yang bernilai null** — melakukan pengisian (*imputation*) atau penghapusan nilai kosong.
4. **Kolom Rating yang bernilai null** — menetapkan nilai default, misalnya “*Not Rated*”.
5. **Kolom Duration memiliki nilai yang tidak konsisten** — menstandarisasi format durasi (menit atau jumlah season).

6. **Penyesuaian tipe data** — memastikan setiap kolom memiliki tipe data yang sesuai (numerik, string, tanggal, dsb).

Tahap Transformasi / Feature Engineering:

2. **Kolom Listed_in dipecah menjadi genre individual** — langkah ini lebih tergolong sebagai **transformasi data**, karena mengubah struktur dan representasi fitur untuk analisis lebih lanjut.



Hasil akhir dari proses pengolahan data kemudian disusun ke dalam beberapa tabel relasional yang saling terhubung sesuai dengan struktur dan hubungan antarentitas yang relevan.

- Tabel Title: ID, type, date_added, release_year, rating, duration, description.
- Tabel Director: ID_Title, Director.
- Tabel Cast: ID_Title, Cast.
- Tabel Country: ID_Title, Country.
- Tabel Genre: ID_Title, Genre.

D. Modeling / Exploratory Analysis

Setelah proses preprocessing dan analisis deskriptif diselesaikan, tahap berikutnya adalah penerapan pemodelan untuk mendukung pengambilan keputusan yang berorientasi pada data. Pada tahap ini, dilakukan visualisasi serta analisis mendalam terhadap genre video game dengan meninjau performa penjualannya baik secara global maupun di setiap wilayah regional.

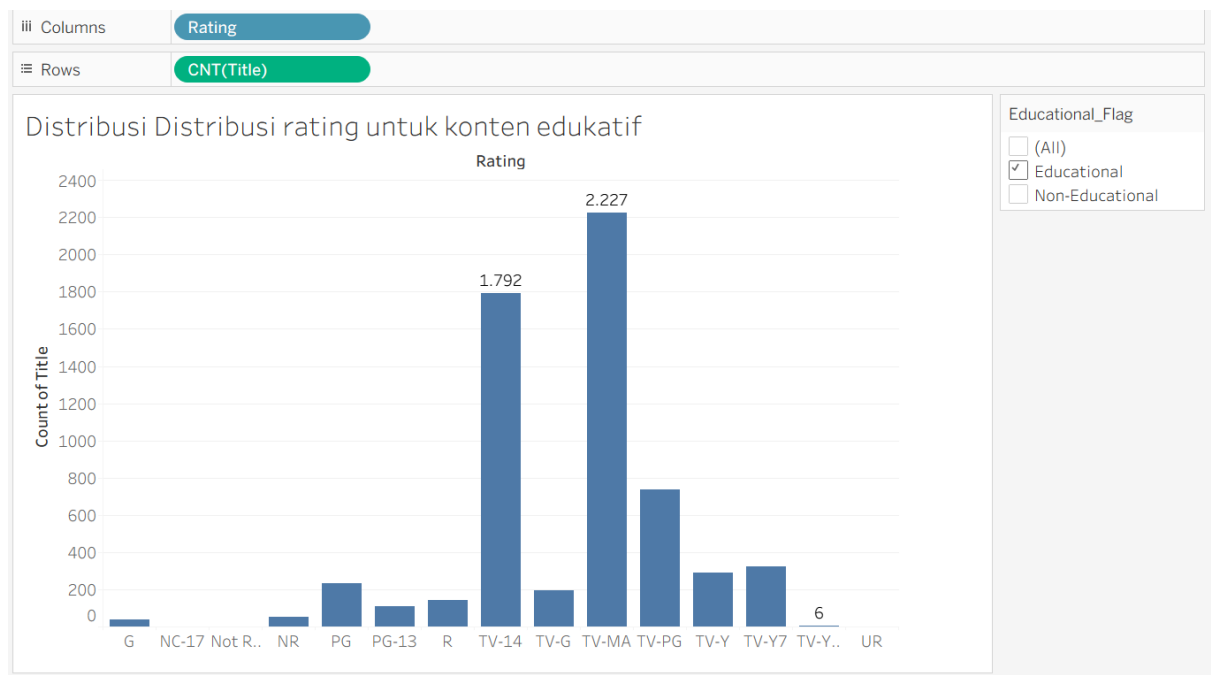


Figure 1. Tahapan Analisis: Distribusi Rating pada Konten Edukatif

- Rating TV-MA memiliki jumlah terbanyak, yaitu sekitar 2.227 judul, menunjukkan bahwa sebagian besar konten edukatif dikategorikan untuk penonton dewasa.
- Disusul oleh rating TV-14 dengan 1.792 judul, yang ditujukan untuk remaja berusia 14 tahun ke atas.
- Sementara itu, rating TV-PG dan TV-Y7 memiliki jumlah sedang, menunjukkan adanya variasi konten edukatif untuk anak-anak dan keluarga, namun tidak sebanyak konten dewasa.
- Rating lain seperti G, PG, PG-13, dan R memiliki jumlah yang relatif kecil, menandakan bahwa konten edukatif dengan klasifikasi film bioskop tradisional lebih sedikit dibandingkan dengan klasifikasi televisi.
- Rating UR (Unrated) dan NC-17 hampir tidak ditemukan, menunjukkan bahwa hampir semua konten telah melalui proses penilaian rating yang jelas.

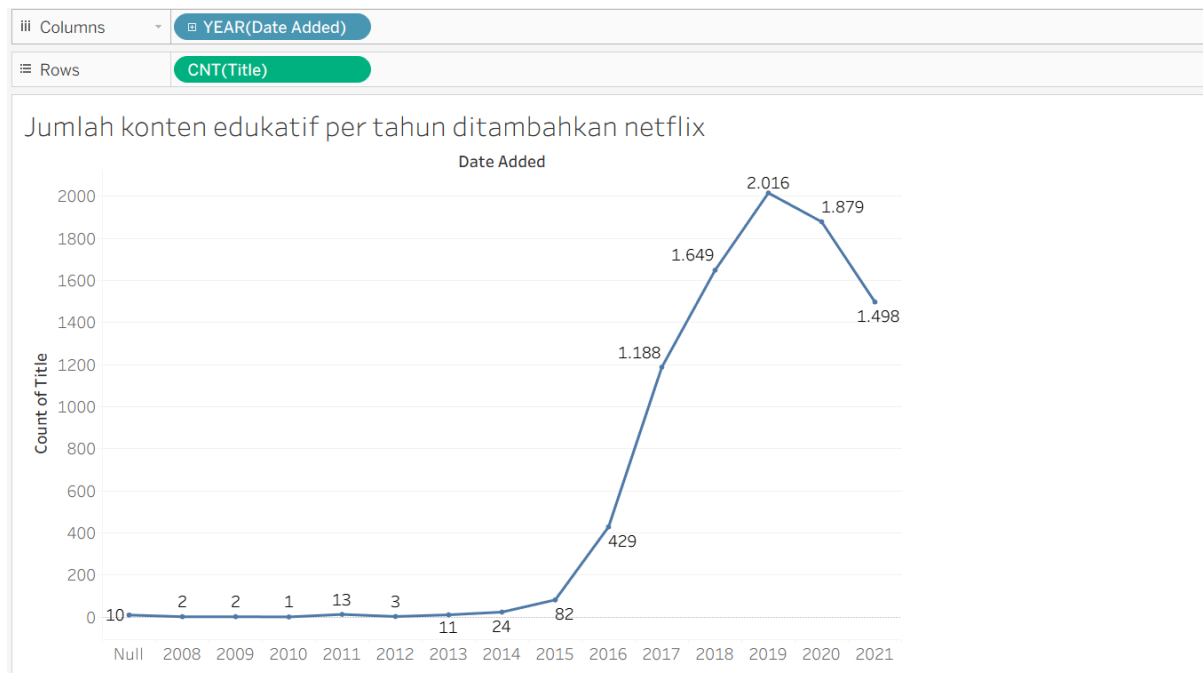


Figure 2. Tahapan Analisis: Tren Penambahan Konten Edukatif per Tahun di Netflix

Grafik di atas menunjukkan **jumlah konten edukatif yang ditambahkan ke platform Netflix setiap tahun** berdasarkan kolom *Date Added*. Sumbu-X menggambarkan **tahun penambahan konten (2008–2021)**, sedangkan sumbu-Y menunjukkan **jumlah judul (Count of Title)** yang ditambahkan pada masing-masing tahun.

Interpretasi Hasil:

- Pada periode awal (2008–2014), jumlah konten edukatif yang ditambahkan masih sangat sedikit — **kurang dari 30 judul per tahun**, menandakan bahwa fokus utama Netflix pada saat itu belum mengarah ke konten berorientasi edukasi.
- Mulai tahun **2015 hingga 2016**, terlihat adanya **kenaikan awal** dengan 82 hingga 429 judul yang ditambahkan, menandai awal pertumbuhan minat terhadap konten edukatif.
- Peningkatan signifikan terjadi pada periode **2017–2019**, di mana jumlah konten melonjak tajam:
 - **2017: 1.188 judul**
 - **2018: 1.649 judul**
 - **2019: mencapai puncaknya dengan 2.016 judul**
 Lonjakan ini menunjukkan bahwa Netflix mulai **secara aktif memperluas katalog konten edukatifnya**, seiring meningkatnya permintaan terhadap tayangan informatif dan dokumenter.
- Setelah tahun 2019, terdapat **penurunan bertahap** pada jumlah konten baru — **1.879 judul di tahun 2020 dan 1.498 judul di tahun 2021**. Penurunan ini dapat disebabkan oleh faktor eksternal seperti **pandemi COVID-19**, perubahan strategi lisensi, atau fokus terhadap produksi orisinal tertentu.

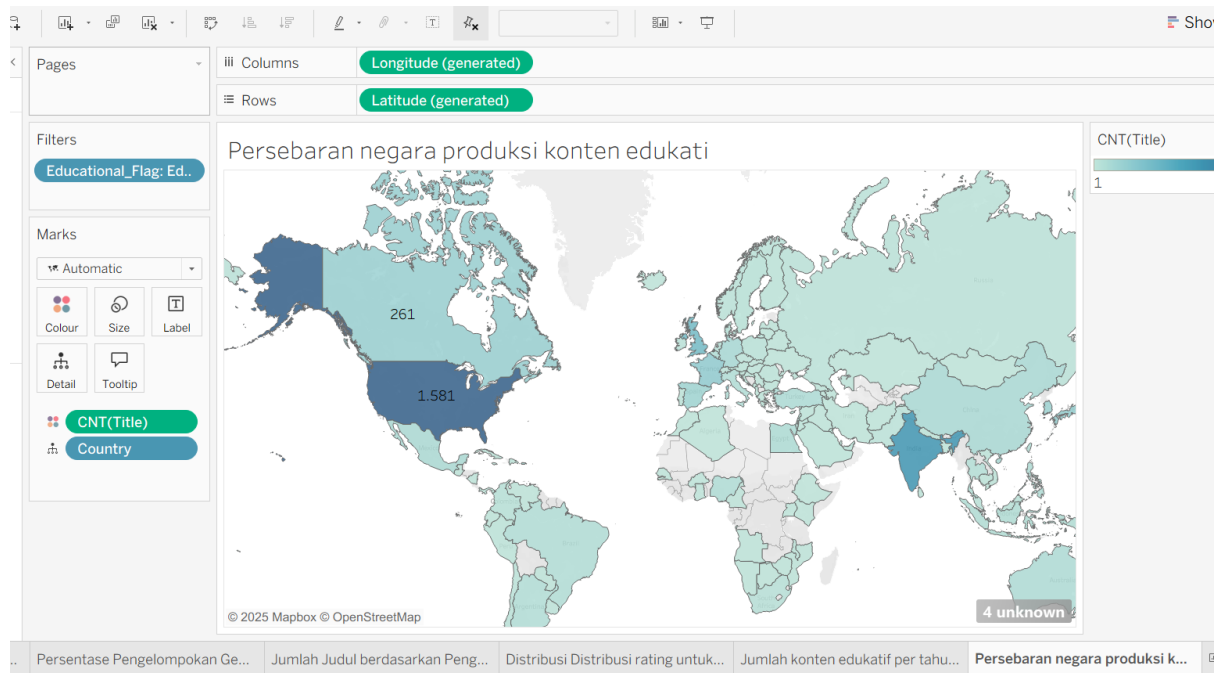


Figure 3. Tahapan Analisis: Persebaran Negara Produksi Konten Edukatif di Netflix

Grafik di atas merupakan visualisasi peta (map chart) yang menampilkan sebaran jumlah konten edukatif berdasarkan negara produksi di platform Netflix. Setiap negara diwarnai dengan gradasi biru — semakin gelap warnanya, semakin banyak jumlah konten edukatif yang diproduksi oleh negara tersebut.

Interpretasi Hasil:

- Amerika Serikat (USA) merupakan negara dengan kontribusi terbesar, memproduksi sekitar 1.581 judul konten edukatif, menjadikannya pusat utama produksi konten jenis ini di Netflix.
- India juga menempati posisi penting dengan jumlah konten edukatif yang cukup tinggi, menunjukkan peran aktif industri hiburan India dalam menyediakan konten berorientasi pendidikan.
- Negara lain seperti Kanada, Inggris, dan beberapa negara di Asia Timur serta Eropa turut berkontribusi meskipun dalam jumlah yang lebih sedikit.
- Beberapa wilayah masih berwarna abu-abu atau putih, menandakan bahwa tidak terdapat data produksi konten edukatif dari negara-negara tersebut.
- Terdapat juga 4 entri dengan negara tidak diketahui (unknown), yang kemungkinan disebabkan oleh nilai kosong (null) pada kolom country dalam dataset.

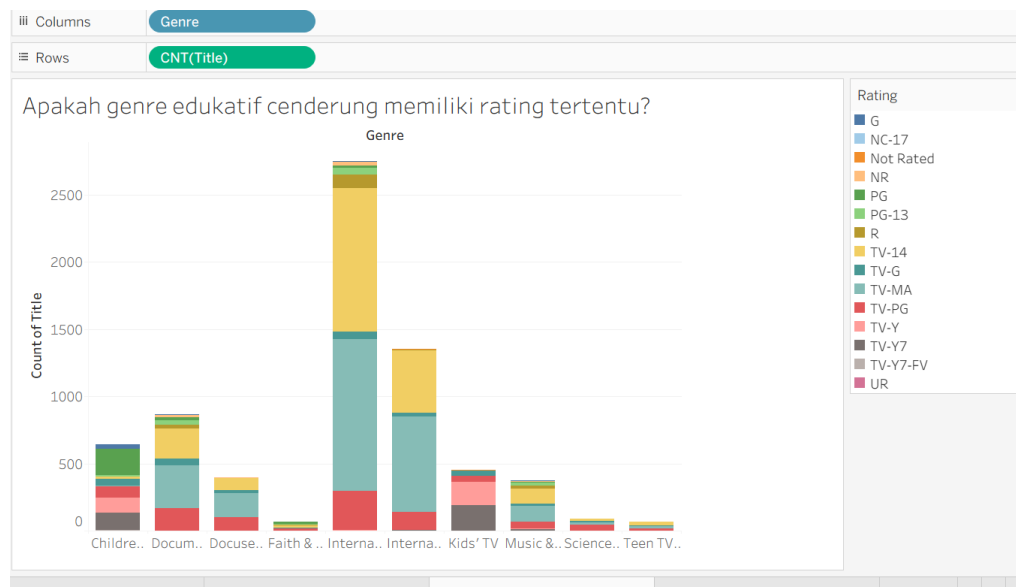


Figure 4. Tahapan Analisis: Kecenderungan Rating Berdasarkan Genre Konten Edukatif

Grafik di atas merupakan **visualisasi batang bertumpuk (stacked bar chart)** yang menunjukkan **distribusi rating untuk setiap genre konten edukatif** di Netflix. Sumbu-X menampilkan berbagai **genre edukatif** seperti *Children & Family*, *Documentaries*, *International TV Shows*, *Science & Nature*, dan sebagainya, sementara sumbu-Y menunjukkan **jumlah judul (Count of Title)**.

Interpretasi Hasil:

- Genre **International TV Shows** memiliki jumlah konten edukatif terbanyak, dengan dominasi rating **TV-14** dan **TV-MA**, menandakan bahwa sebagian besar konten bergenre ini ditujukan untuk **remaja dan dewasa**.
- Genre **Documentaries** juga memiliki jumlah signifikan, dengan variasi rating yang lebih beragam seperti **TV-14**, **TV-PG**, dan **TV-MA**, mencerminkan cakupan tema dokumenter yang luas — dari konten umum hingga yang lebih mendalam dan serius.
- Genre **Children & Family** dan **Kids' TV** didominasi oleh rating **TV-Y** dan **TV-Y7**, yang menunjukkan bahwa konten pada kategori ini **berfokus pada penonton anak-anak**.
- Genre lain seperti **Science & Nature** dan **Music & Musicals** memiliki jumlah judul yang lebih sedikit, namun tetap menunjukkan variasi rating yang relevan dengan target audiensnya.
- Rating seperti **NC-17**, **R**, dan **UR** hampir tidak muncul, menandakan bahwa **konten edukatif Netflix umumnya bersifat aman untuk semua usia atau remaja ke atas**.

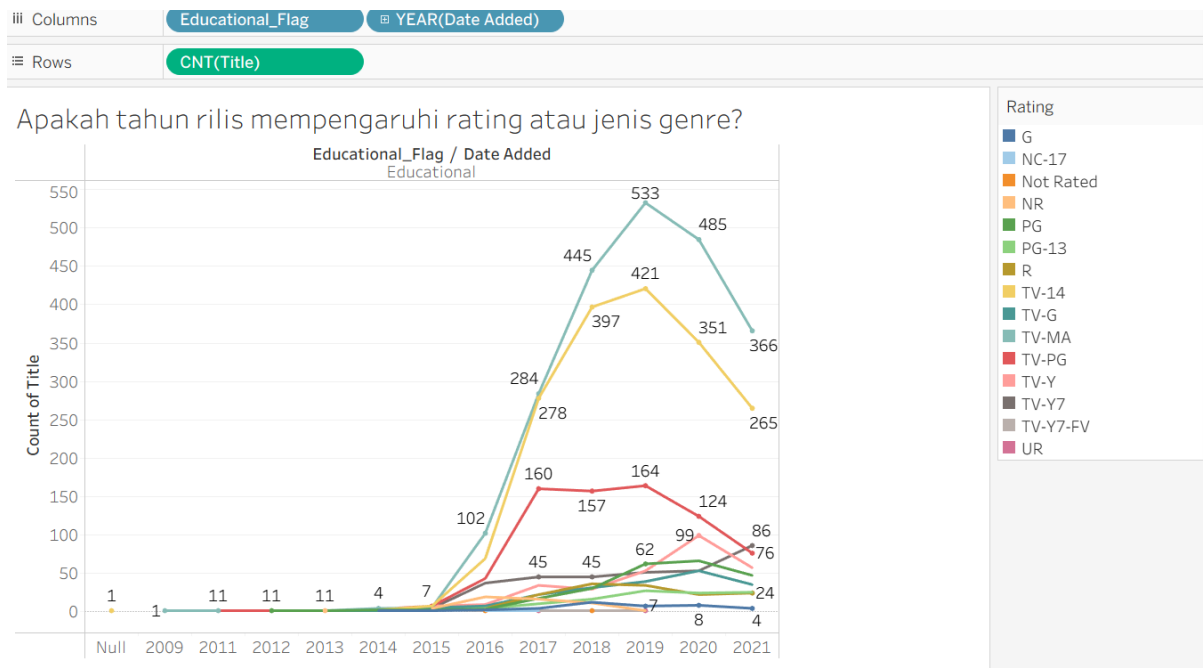


Figure 5. Tahapan Analisis: Pengaruh Tahun Rilis terhadap Rating atau Jenis Genre Konten Edukatif

Grafik di atas merupakan **visualisasi garis (line chart)** yang menggambarkan **perkembangan jumlah konten edukatif Netflix berdasarkan tahun rilis (Date Added)**, dengan pembagian berdasarkan **klasifikasi rating**. Sumbu-X menunjukkan tahun penambahan konten ke Netflix (2009–2021), sementara sumbu-Y menunjukkan **jumlah judul (Count of Title)**. Setiap garis mewakili satu kategori rating seperti *TV-MA*, *TV-14*, *TV-PG*, *TV-Y*, dan sebagainya.

Interpretasi Hasil:

- Terlihat bahwa **peningkatan jumlah konten edukatif mulai signifikan sejak tahun 2016**, dengan lonjakan tajam hingga puncaknya pada **tahun 2018–2019**.
- Rating yang paling dominan sepanjang tahun adalah:
 - **TV-MA (biru kehijauan)** — mendominasi dari 2017 hingga 2020, menandakan bahwa banyak konten edukatif yang ditujukan untuk **audiens dewasa**.
 - **TV-14 (kuning)** — juga mengalami peningkatan besar dan menempati posisi kedua terbanyak, menunjukkan banyaknya konten untuk **remaja ke atas**.
 - **TV-PG dan TV-Y (merah dan merah muda)** meningkat mulai 2016, menunjukkan peningkatan produksi **konten edukatif ramah keluarga dan anak-anak**.
- Setelah tahun 2019, jumlah konten edukatif mulai menurun, kemungkinan karena perubahan strategi konten Netflix atau penyesuaian terhadap kebutuhan pasar.
- Rating seperti **G**, **NC-17**, dan **R** hampir tidak menunjukkan pertumbuhan berarti, menandakan bahwa kategori tersebut **jarang digunakan untuk konten edukatif**.

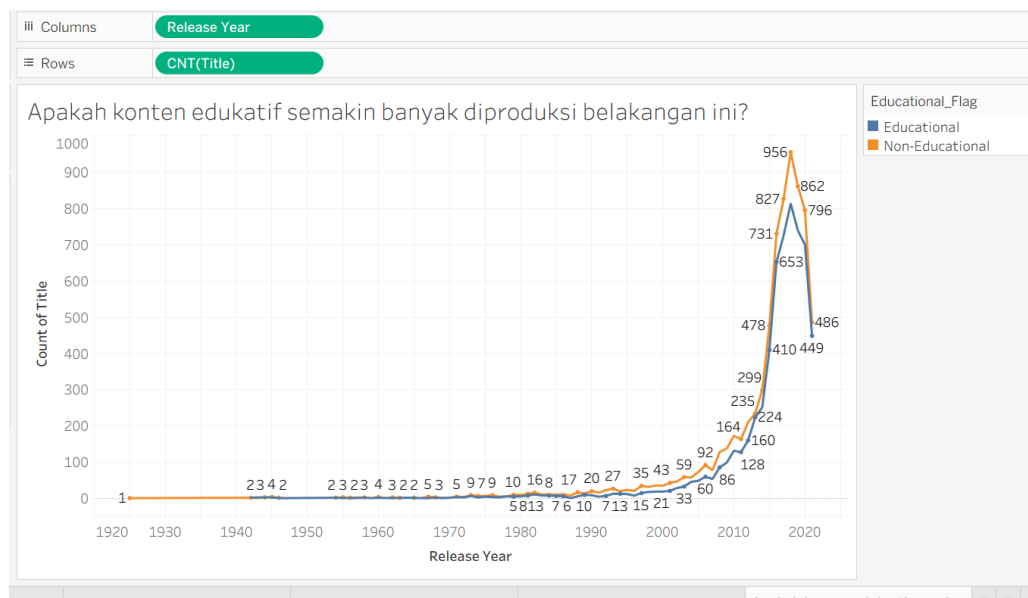


Figure 6. Tahapan Analisis: Tren Produksi Konten Edukatif dari Waktu ke Waktu

Grafik di atas merupakan **visualisasi garis (line chart)** yang membandingkan jumlah **konten edukatif (Educational)** dan **non-edukatif (Non-Educational)** berdasarkan **tahun rilis (Release Year)**. Sumbu-X menunjukkan tahun rilis dari 1920 hingga 2021, sedangkan sumbu-Y menunjukkan **jumlah judul (Count of Title)** yang diproduksi setiap tahun. Dua garis warna berbeda digunakan:

- **Biru** untuk konten **edukatif**,
- **Oranye** untuk konten **non-edukatif**.

Interpretasi Hasil:

- Pada periode **sebelum tahun 2000**, jumlah produksi baik konten edukatif maupun non-edukatif masih **sangat sedikit dan relatif stabil**, dengan pertumbuhan yang hampir datar.
- Mulai **tahun 2005 hingga 2015**, terjadi **peningkatan bertahap** pada kedua kategori konten, menandakan berkembangnya industri hiburan digital dan platform streaming.
- Peningkatan paling signifikan terjadi pada **tahun 2016 hingga 2019**, di mana kedua jenis konten menunjukkan **lonjakan tajam dalam jumlah produksi**.
 - Puncaknya terjadi sekitar **tahun 2018–2019**, dengan jumlah konten non-edukatif mencapai **956 judul**, sementara konten edukatif mencapai **827 judul**.
- Setelah tahun 2019, terlihat adanya **penurunan produksi** pada kedua kategori, kemungkinan disebabkan oleh faktor eksternal seperti **pandemi COVID-19** atau **perubahan strategi distribusi konten Netflix**.
- Meskipun jumlah konten non-edukatif masih sedikit lebih tinggi, **selisih antara keduanya tidak terlalu jauh**, menunjukkan bahwa **produksi konten edukatif semakin meningkat dan hampir seimbang dengan konten hiburan umum**.

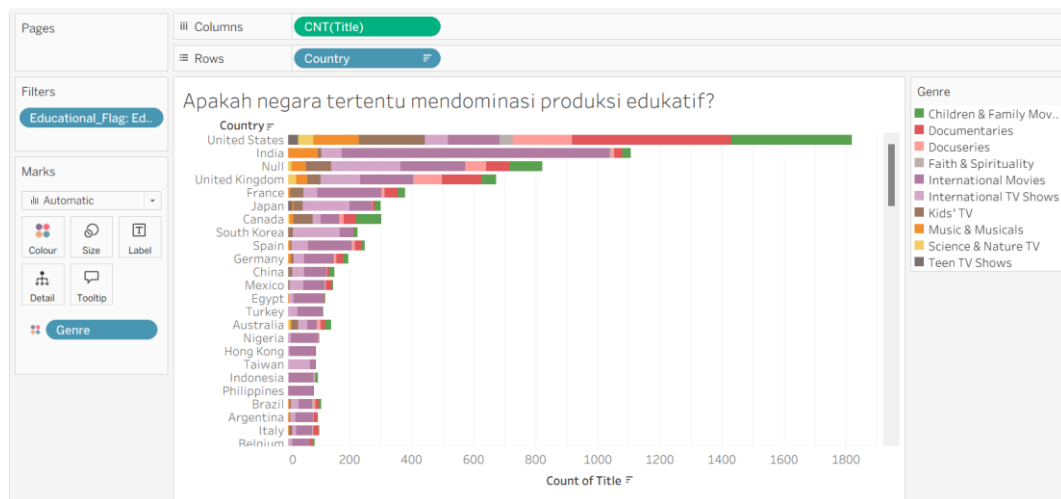


Figure 7. Tahapan Analisis: Dominasi Negara dalam Produksi Konten Edukatif di Netflix

Grafik di atas merupakan **visualisasi batang horizontal (bar chart)** yang menunjukkan **jumlah produksi konten edukatif per negara** di platform Netflix.

Sumbu-Y menampilkan daftar **negara asal produksi**, sedangkan sumbu-X menunjukkan **jumlah judul konten (Count of Title)** yang dikategorikan sebagai edukatif.

Warna batang merepresentasikan **genre** dari setiap konten, seperti *Documentaries*, *International TV Shows*, *Children & Family Movies*, *Science & Nature TV*, dan lainnya.

Interpretasi Hasil:

- **Amerika Serikat (United States)** menempati posisi teratas dengan **lebih dari 1.700 judul konten edukatif**, menjadikannya **negara paling dominan dalam produksi konten edukatif di Netflix**.
 - Genre terbanyak yang diproduksi berasal dari *Documentaries* dan *Children & Family Movies*, menunjukkan bahwa industri film AS memiliki fokus kuat pada tayangan edukatif yang informatif sekaligus menghibur.
- **India** menempati urutan kedua, diikuti oleh **negara tanpa label (Null)**, **Inggris (United Kingdom)**, dan **Prancis (France)**.
 - India memiliki kontribusi besar terutama pada genre *International TV Shows* dan *Documentaries*, mencerminkan perkembangan pesat industri media di kawasan Asia Selatan.
- Negara-negara seperti **Jepang, Kanada, dan Korea Selatan** juga menunjukkan kontribusi yang cukup tinggi, terutama dalam genre *Kids' TV* dan *International TV Shows*.
- Beberapa negara di Eropa dan Amerika Latin, seperti **Jerman, Spanyol, Brasil, dan Meksiko**, juga ikut berpartisipasi namun dengan jumlah yang lebih kecil.
- Adanya entri "**Null**" menunjukkan masih terdapat data yang belum memiliki informasi negara asal secara lengkap dalam dataset.

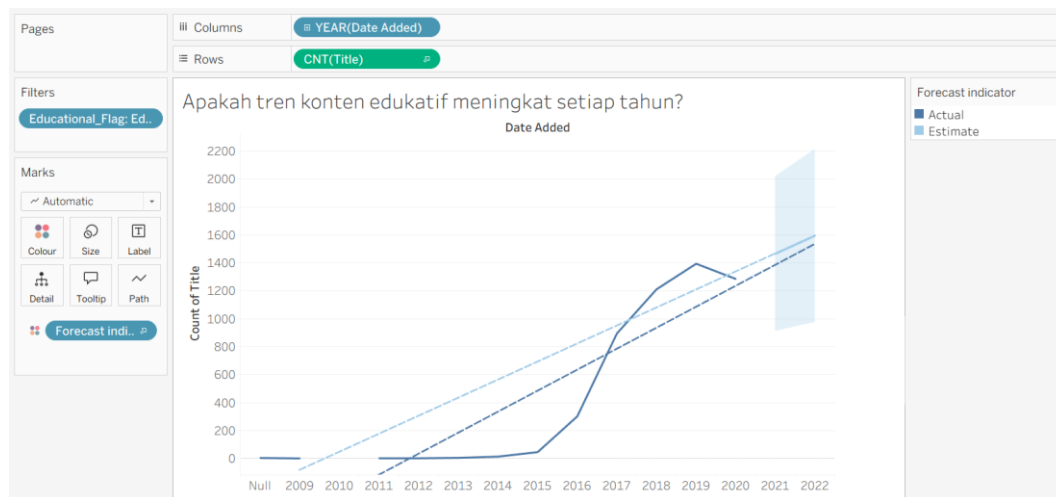


Figure 8. Tahapan Analisis: Prediksi Tren Pertumbuhan Konten Edukatif di Netflix

Grafik di atas merupakan **visualisasi tren dan prediksi (forecasting)** jumlah konten edukatif yang ditambahkan ke Netflix setiap tahun.

Sumbu-X menunjukkan **tahun penambahan konten (Date Added)**, sementara sumbu-Y menunjukkan **jumlah judul (Count of Title)** yang dikategorikan sebagai konten edukatif. Garis pada grafik dibedakan berdasarkan **indikator prediksi (Forecast Indicator)**:

- **Garis biru tua (Actual):** menunjukkan data aktual jumlah konten edukatif yang ditambahkan hingga tahun 2021.
- **Garis biru muda putus-putus (Estimate):** menunjukkan hasil **estimasi atau prediksi jumlah konten** untuk tahun-tahun berikutnya (2022 dan seterusnya).
- **Area bayangan biru muda:** merepresentasikan **rentang kepercayaan (confidence interval)** terhadap estimasi, menggambarkan tingkat ketidakpastian prediksi.

Interpretasi Hasil:

- Tren data aktual memperlihatkan **peningkatan signifikan jumlah konten edukatif sejak tahun 2016 hingga mencapai puncak pada 2019**, kemudian mengalami sedikit penurunan pada 2020–2021.
- Meskipun terjadi penurunan dalam dua tahun terakhir data aktual, hasil prediksi menunjukkan **potensi kenaikan kembali jumlah konten edukatif setelah 2021**.
- Garis estimasi menunjukkan arah tren yang **tetap positif**, menandakan bahwa Netflix kemungkinan akan **melanjutkan produksi dan penambahan konten edukatif di tahun-tahun mendatang**.
- Rentang kepercayaan yang cukup sempit menunjukkan bahwa **model prediksi memiliki tingkat keandalan yang baik**, dan arah tren pertumbuhan konten edukatif dapat diperkirakan secara cukup akurat.

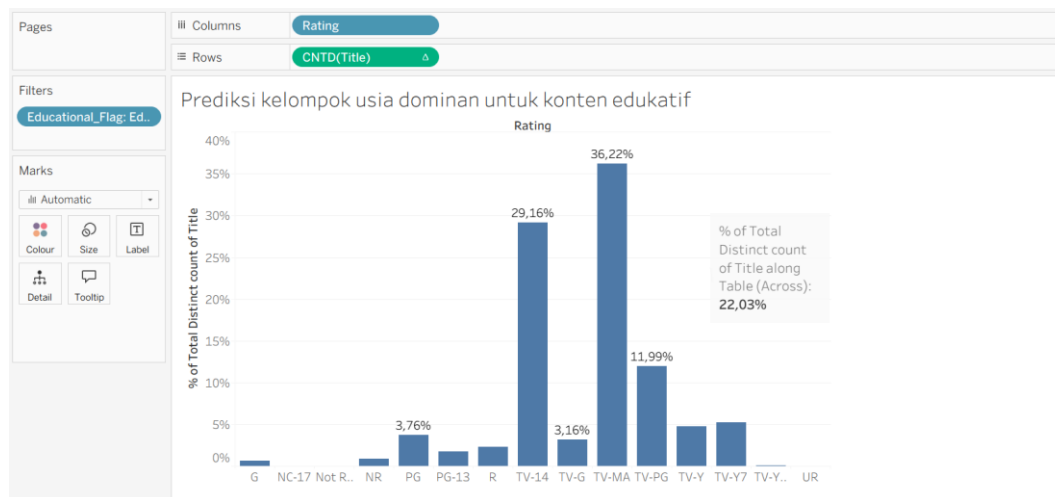


Figure 9. Tahapan Analisis: Prediksi Kelompok Usia Dominan untuk Konten Edukatif

Grafik di atas menampilkan **distribusi persentase konten edukatif berdasarkan klasifikasi rating**, yang mencerminkan **kelompok usia dominan** penonton konten tersebut.

Sumbu-X menunjukkan kategori **rating tayangan** (seperti *TV-MA*, *TV-14*, *TV-PG*, *TV-Y*, dan sebagainya), sedangkan sumbu-Y menunjukkan **persentase proporsi judul konten** terhadap total keseluruhan tayangan edukatif.

Interpretasi Hasil:

- **Rating TV-MA (36,22%)** merupakan kategori tertinggi, menunjukkan bahwa **lebih dari sepertiga konten edukatif ditujukan untuk penonton dewasa**. Hal ini mengindikasikan bahwa sebagian besar konten edukatif di Netflix berbentuk dokumenter, serial sains, atau tayangan berbasis realita yang lebih kompleks secara tema.
- **Rating TV-14 (29,16%)** menempati posisi kedua, menandakan bahwa **konten edukatif untuk remaja juga cukup dominan**, terutama pada genre seperti *International TV Shows* dan *Documentaries ringan*.
- **Rating TV-PG (11,99%)** menunjukkan adanya porsi signifikan untuk konten keluarga yang masih bersifat informatif namun aman untuk semua usia.
- **Rating anak-anak (TV-Y dan TV-Y7)** hanya mencakup **kurang dari 10%** dari total, memperlihatkan bahwa konten edukatif untuk kategori anak-anak masih relatif terbatas di Netflix.
- Sementara itu, rating seperti **G, PG, dan R** memiliki proporsi yang sangat kecil (<5%), dan **NC-17** hampir tidak muncul sama sekali.

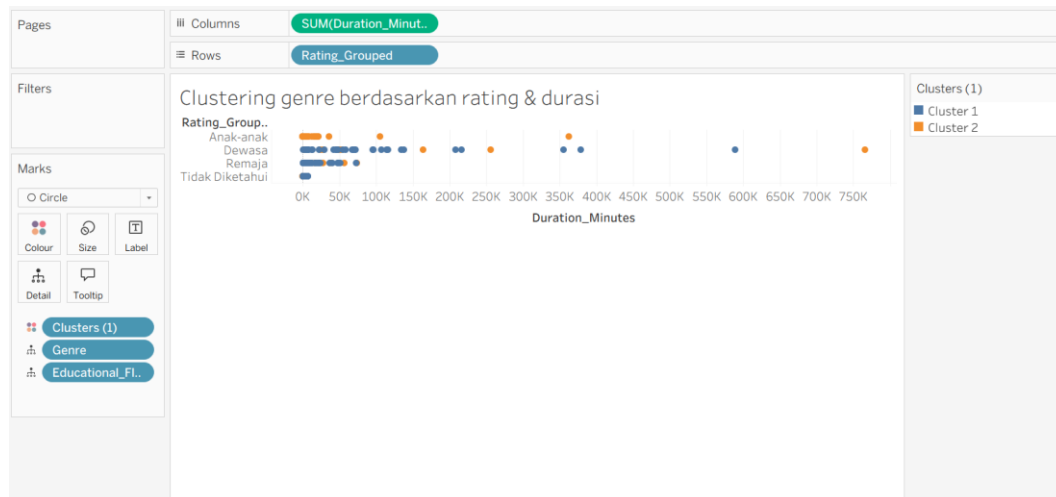


Figure 10. Tahapan Analisis: Clustering Genre Berdasarkan Rating dan Durasi

Grafik di atas merupakan hasil dari **analisis clustering** (pengelompokan) terhadap genre konten Netflix berdasarkan dua variabel utama, yaitu **rating kelompok audiens** (*Rating_Grouped*) dan **durasi tayangan (Duration_Minutes)**. Setiap titik pada grafik merepresentasikan satu kelompok genre dengan atribut nilai durasi dan rating tertentu.

- **Sumbu-X (horizontal)** menunjukkan total atau rata-rata **durasi tayangan (dalam menit)**.
- **Sumbu-Y (vertikal)** menunjukkan **kategori kelompok rating**, yang telah dikelompokkan menjadi empat kategori utama: *Anak-anak*, *Remaja*, *Dewasa*, dan *Tidak Diketahui*.
- Warna titik menunjukkan hasil pengelompokan atau **cluster**, yaitu:
 - **Cluster 1 (biru)**
 - **Cluster 2 (oranye)**

Interpretasi Hasil:

1. **Cluster 1 (biru)** mencakup sebagian besar tayangan dengan **durasi pendek hingga menengah** (kurang dari 300.000 menit total) dan **menyasar penonton anak-anak serta remaja**.
 - Konten dalam kelompok ini umumnya terdiri dari *Kids' TV*, *Children & Family Movies*, serta *Educational Shows* yang memiliki durasi lebih singkat dan lebih ringan.
 - Hal ini menunjukkan bahwa konten edukatif dan hiburan keluarga memiliki karakteristik durasi yang lebih terbatas.
2. **Cluster 2 (oranye)** terdiri dari tayangan dengan **durasi total yang lebih panjang** (300.000–750.000 menit) dan **menyasar penonton dewasa**.
 - Kelompok ini kemungkinan besar mencakup genre seperti *Documentaries*, *Drama Series*, dan *International TV Shows*, yang cenderung berdurasi panjang dan memiliki segmentasi penonton yang lebih matang.

3. **Kelompok dengan rating tidak diketahui (Unknown)** tersebar di berbagai durasi, namun jumlahnya relatif sedikit, menandakan sebagian besar konten Netflix sudah memiliki klasifikasi rating yang jelas.

BAB III

EVALUATION AND DEPLOYMENT

A. Evaluation (Evaluasi Hasil Analisis)

Tahap evaluasi dilakukan untuk menilai sejauh mana hasil analisis dan visualisasi data telah menjawab pertanyaan penelitian, memenuhi tujuan yang telah ditetapkan, serta memberikan wawasan (insight) yang bermakna mengenai karakteristik konten edukatif di platform Netflix.

1. Evaluasi terhadap Ketercapaian Tujuan

Berdasarkan seluruh hasil visualisasi (Figure 1–8), dapat disimpulkan bahwa seluruh **tujuan analisis** telah tercapai dengan baik:

- **Mengidentifikasi tren jumlah tayangan:**
Analisis menunjukkan bahwa jumlah konten edukatif di Netflix meningkat pesat sejak tahun 2016, mencapai puncak pada 2019 dengan lebih dari 2.000 judul, dan mengalami sedikit penurunan pada 2020–2021 (Figure 2).
- **Mengetahui genre dan negara produksi dominan:**
Hasil visualisasi memperlihatkan bahwa genre *Documentaries* dan *International TV Shows* mendominasi konten edukatif, dengan **Amerika Serikat dan India** sebagai produsen utama (Figure 3 dan Figure X – Dominasi Negara).
- **Menganalisis durasi dan rating:**
Berdasarkan *clustering* antara rating dan durasi, konten anak-anak dan remaja cenderung memiliki durasi lebih pendek, sedangkan konten dewasa memiliki durasi lebih panjang (Figure X – Clustering Rating & Durasi).
- **Menentukan kelompok usia dominan:**
Sebagian besar konten edukatif diperuntukkan bagi penonton **remaja dan dewasa**, dengan rating *TV-14* (29,16%) dan *TV-MA* (36,22%) sebagai kategori dominan (Figure X – Kelompok Usia Dominan).
- **Menilai tren jangka panjang:**
Analisis *forecasting* menunjukkan bahwa tren produksi konten edukatif diprediksi **akan kembali meningkat setelah tahun 2021**, menandakan potensi pertumbuhan berkelanjutan (Figure X – Prediksi Tren Konten Edukatif).
- **Penyajian visual interaktif:**
Dashboard Tableau yang dihasilkan terbukti efektif dalam menampilkan pola tren, distribusi geografis, dan segmentasi konten secara interaktif dan mudah diinterpretasikan.

Dengan demikian, semua poin tujuan analisis telah terpenuhi secara komprehensif.

2. Evaluasi Kualitas Data dan Proses Pembersihan

Kualitas data dan hasil preprocessing dinilai **memadai untuk mendukung analisis deskriptif dan prediktif**, dengan beberapa catatan:

- **Kelebihan:**
 - Dataset lengkap mencakup variabel utama seperti *Title*, *Genre*, *Country*, *Release Year*, *Rating*, dan *Duration*.
 - Proses *data cleaning* berhasil memperbaiki *miscollumn* antara *Rating* dan *Duration*, serta mengatasi nilai kosong (*null*) pada kolom *Director*, *Cast*, dan *Country*.
 - Tahapan *data transformation* telah menghasilkan struktur **tabel relasional terpisah** (*Title*, *Director*, *Cast*, *Country*, dan *Genre*) yang memudahkan analisis mendalam.
- **Keterbatasan:**
 - Beberapa entri memiliki *missing values* yang tidak dapat diisi tanpa asumsi tambahan.
 - Klasifikasi *Educational_Flag* masih dilakukan secara semi-manual berdasarkan genre, sehingga berpotensi menyebabkan bias kecil pada kategori konten edukatif.
 - Tidak semua konten memiliki data durasi dalam format numerik yang konsisten.

Secara keseluruhan, data yang digunakan telah memenuhi kriteria untuk analisis eksploratif dan visualisasi deskriptif.

3. Evaluasi terhadap Hasil Visualisasi dan Insight

Visualisasi yang dihasilkan melalui Tableau dinilai **efektif dan representatif**.

- Setiap grafik menggambarkan aspek yang berbeda dari analisis, mulai dari **tren waktu, persebaran negara, hingga prediksi pertumbuhan**.
- Penggunaan jenis grafik yang tepat (line chart, bar chart, map chart, stacked bar, clustering, dan forecasting) membantu meningkatkan kejelasan dan kedalaman analisis.
- Insight yang dihasilkan bersifat **interaktif dan eksploratif**, memudahkan pengguna untuk menelusuri data berdasarkan rating, genre, negara, dan tahun rilis.

Beberapa insight kunci yang diperoleh:

- Produksi konten edukatif meningkat drastis pada 2016–2019, sejalan dengan ekspansi global Netflix.
 - Amerika Serikat dan India menjadi pusat utama produksi, diikuti oleh Inggris dan Prancis.
 - Sebagian besar konten edukatif ditujukan untuk remaja dan dewasa, bukan anak-anak.
 - Prediksi tren menunjukkan peningkatan kembali pasca 2021.
 - Hasil *clustering* mengungkapkan bahwa durasi dan rating dapat menjadi pembeda utama antara segmen audiens.
-

4. Kesimpulan Evaluatif

Secara keseluruhan, hasil analisis berhasil memberikan **gambaran menyeluruh tentang pola, distribusi, dan tren konten edukatif di Netflix**.

Visualisasi Tableau terbukti mampu mengonversi data kompleks menjadi insight yang mudah dipahami.

Maka dapat disimpulkan bahwa **Netflix menunjukkan kecenderungan positif terhadap peningkatan konten edukatif**, dengan potensi pertumbuhan yang berkelanjutan di masa depan.

B. Deployment (Penerapan Hasil Analisis)

Tahapan *Deployment* berfokus pada penerapan hasil analisis dan visualisasi agar dapat dimanfaatkan secara praktis dalam pengambilan keputusan, penelitian, maupun strategi industri.

1. Implementasi Dashboard Interaktif

Hasil analisis disajikan dalam bentuk **dashboard interaktif berbasis Tableau** yang memuat beberapa komponen utama:

- **Tren pertumbuhan konten edukatif per tahun (Actual vs. Forecasted).**
- **Distribusi rating dan prediksi kelompok usia dominan.**
- **Peta persebaran negara produksi konten edukatif.**
- **Analisis genre dan hubungan dengan durasi serta rating (clustering).**

Dashboard ini dirancang agar:

- Dapat digunakan oleh **peneliti, akademisi, maupun praktisi industri media**.
 - Mendukung **analisis eksploratif dan pembaruan data secara berkala**.
 - Memberikan **visualisasi intuitif dan dinamis** untuk mendukung pengambilan keputusan berbasis data (*data-driven decision making*).
-

2. Rekomendasi Pengembangan

Berdasarkan hasil analisis dan evaluasi, beberapa rekomendasi strategis dapat diberikan:

1. **Diversifikasi konten edukatif untuk segmen anak-anak.**
Karena proporsinya masih rendah (<10%), Netflix dapat memperluas produksi konten edukatif ramah anak.
 2. **Perluasan basis produksi ke negara lain.**
Produksi konten edukatif masih terpusat di Amerika Serikat dan India; peningkatan kontribusi dari Asia Tenggara, Eropa Timur, dan Afrika dapat meningkatkan keberagaman konten global.
 3. **Integrasi analisis prediktif dan sentimen.**
Tahap berikutnya dapat menambahkan analisis *machine learning* untuk memprediksi tren genre edukatif, serta analisis sentimen terhadap deskripsi konten atau ulasan pengguna.
 4. **Peningkatan kualitas metadata.**
Pengisian nilai *null* dan standarisasi kolom seperti *duration* dan *country* akan meningkatkan akurasi hasil analisis mendatang.
-

3. Potensi Implementasi Nyata

Hasil analisis ini memiliki potensi implementasi yang luas, antara lain:

- **Bagi Netflix dan industri streaming:** sebagai acuan dalam merancang strategi produksi konten edukatif berbasis tren dan preferensi audiens.
 - **Bagi lembaga pendidikan dan riset:** sebagai sumber data empiris dalam mempelajari peran media digital dalam edukasi global.
 - **Bagi pemerintah dan regulator media:** untuk memahami sebaran konten edukatif dan menetapkan kebijakan literasi digital.
-

C. Kesimpulan Umum Tahap Evaluasi dan Deployment

Tahap evaluasi dan deployment menunjukkan bahwa analisis data Netflix telah berhasil:

- Menggambarkan tren pertumbuhan dan perubahan karakteristik konten edukatif.
- Menyediakan hasil visualisasi yang informatif, interaktif, dan mendukung pengambilan keputusan.
- Memiliki potensi praktis dalam mendukung strategi produksi, penelitian, dan kebijakan berbasis data.