

DV : StreamScope

Netflix Content Strategy Analyzer: Insights into Global Streaming Trends

Objective

To develop a data-driven system that analyzes Netflix's content catalog using the Kaggle Netflix Movies and TV Shows dataset.

The project aims to uncover trends in Netflix's expansion strategy, the balance between original vs. licensed content, genre popularity, and preferences across global markets using only this dataset.

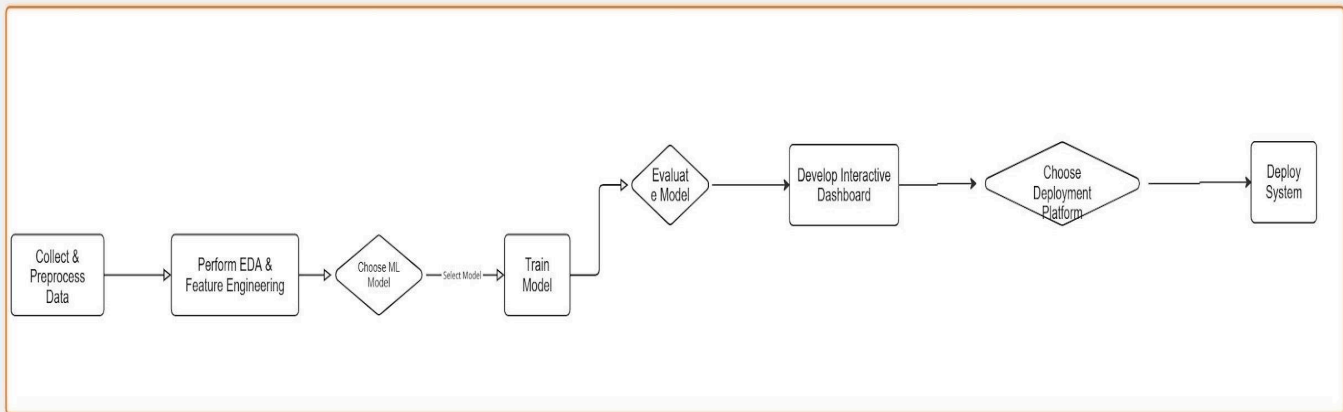
Outcomes

- Understand preprocessing and feature engineering on entertainment datasets.
 - Perform exploratory data analysis (EDA) to identify key trends in Netflix content strategy.
 - Gain experience in clustering and classification techniques for media data.
 - Deliver a functional prototype dashboard showcasing Netflix content insights by genre, region, and time.
-

Dataset

- **Netflix Movies and TV Shows Dataset (Kaggle)**
Contains details of 8,000+ titles including type (Movie/TV Show), director, cast, country, release year, rating, duration, and listed genres.
Link: [Netflix Kaggle Dataset](#)
-

Architectural Diagram



Modules to be Implemented

1. **Data Collection & Preprocessing**
 2. **Exploratory Data Analysis (EDA) & Feature Engineering**
 3. **Modeling & Advanced Analytics**
 4. **Interactive Dashboard & Insights Visualization**
 5. **Evaluation and Final Deployment**
-

Week-wise Module Implementation

Milestone 1: Week 1 & 2 - Requirements & Dataset Preparation

- Define project scope and success metrics.
- Load the Netflix Kaggle dataset.
- Clean the dataset (handle missing values, remove duplicates).
- Normalize categorical features such as genre, rating, and country.

Milestone 2: Week 3 & 4 - EDA & Feature Engineering

- Analyze Netflix content growth over time.

- Visualize the distribution of genres, ratings, and content type.
- Identify country-level content contributions.
- Create derived features such as “Content Length Category” and “Original vs. Licensed” (if derivable from dataset).

Milestone 3: Week 5 & 6 - Modeling & Advanced Analysis

- Use clustering to group Netflix titles by genre, duration, and ratings.
- Classification of content type (Movie vs. TV Show) based on features.
- Analyze key drivers for content availability across countries and genres.
- Use feature importance methods to interpret results.

Milestone 4: Week 7 & 8 - Dashboard, Integration & Deployment

- Develop an interactive dashboard (Streamlit or Tableau) using only the Netflix dataset.
- Provide filters for year, genre, country, and content type.
- Show insights such as top genres per year, content distribution by country, and rating analysis.
- Finalize testing and deploy the dashboard.

Evaluation Criteria

Milestone 1:

- Completion of dataset cleaning and preparation.
- Proper handling of missing and categorical values.

Milestone 2:

- Insightful EDA with clear visualizations (genre trends, growth analysis).
- Logical and interpretable feature engineering.

Milestone 3:

- Effective clustering and classification results.
- Transparent interpretation of findings.

Milestone 4:

- Fully functional interactive dashboard.
- Smooth filtering and insights visualization.
- Well-documented project presentation.

Tools & Tech Stack

Programming Language

- Python: for data preprocessing, analysis, and modeling.

Libraries & Frameworks

- **Data Handling:** pandas, numpy
- **Visualization:** matplotlib, plotly, seaborn
- **Machine Learning:** scikit-learn
- **Dashboard/UI:** Streamlit, Tableau, or Power BI

Development Tools

- IDE: Jupyter Notebook, VS Code
- Version Control: Git with GitHub
- Deployment: Streamlit Cloud, Heroku, or AWS