

# NETFLIX MOVIES AND TV SHOWS CLUSTERING

HRITIK SHARMA  
AlmaBetter, Bangalore

## Abstract:

**Netflix** is an American subscription streaming service and production company. Launched on August 29, 1997, it offers a film and television series library through duration deals as well as its own productions, known as Netflix Originals. Netflix was founded on the aforementioned date by *Reed Hastings* and *Marc Randolph* in Scotts Valley, California.

As of December 31, 2021, Netflix had over 221.8 million subscribers worldwide.

Netflix can be accessed via internet browsers on computers, or via application software installed on smart TVs, set-top boxes connected to television, tablet, computers, smartphone etc.

## 1.Problem Statement

This dataset consists of tv shows and movies available on Netflix as of 2019. The dataset is collected from Flexible which is a third-party Netflix search engine.

In 2018, they released an interesting report which shows that the number of TV shows on Netflix has nearly tripled since 2010. The streaming service's number of movies has decreased by more than 2,000 titles since 2010, while its number of TV shows has nearly tripled. It will be interesting to explore what all other insights can be obtained from the same dataset.

In this project, you are required to do

1. Exploratory Data Analysis

2. Understanding what type content is available in different countries
3. Netflix has increasingly focused on TV rather than movies in recent years.
4. Clustering similar content by matching text-based features

## **2.Data Summary**

- show\_id : Unique ID for every Movie / Tv Show
- type : Identifier - A Movie or TV Show
- title : Title of the Movie / Tv Show
- director : Director of the Movie
- cast : Actors involved in the movie / show
- country : Country where the movie / show was produced
- date\_added : Date it was added on Netflix
- release\_year : Actual Release Year of the movie / show
- rating : TV Rating of the movie / show
- duration : Total Duration - in minutes or number of seasons
- listed\_in : Genre
- description: The Summary description

## **3. Steps involved:**

- **Exploratory Data Analysis**

After loading the dataset we performed eda. This process helped us figure out various aspects and relationships . It gave us a better idea of which feature behaves in which manner.

- **Null values Treatment**

Our dataset contains many null values. We fill them with some other values and we also drop some of them because null values might tend to disturb our model.

- **Standardization of features**

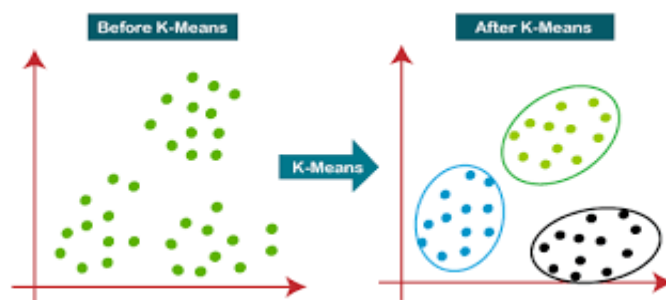
Our main motive through this step was to scale our data into a uniform format that would allow us to utilize the data in a better way while performing fitting and applying different algorithms to it.

The basic goal was to enforce a level of consistency or uniformity to certain practices or operations within the selected environment.

#### 4. Algorithms:

- **K - means clustering:**

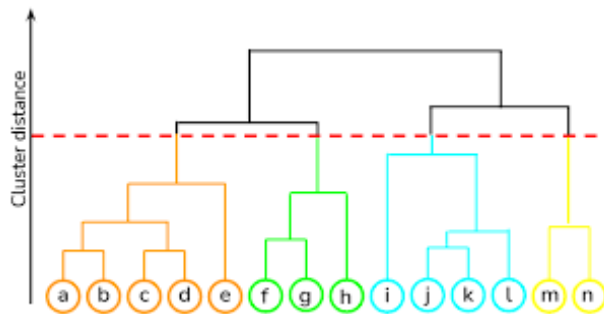
k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.



- **Hierarchical clustering:**

Hierarchical clustering, also known as *hierarchical cluster analysis*, is an algorithm that groups similar objects into groups called *clusters*. The endpoint is a set of clusters, where each cluster is distinct from

each other cluster, and the objects within each cluster are broadly similar to each other.



## 5. Conclusion:

That's it! We reached the end of our exercise.

Starting with loading the data so far we have done EDA , null values treatment, encoding of categorical columns, feature selection and then clustering.

It's clear that Netflix has grown over the years. We can see from the data that the company took certain approaches in their marketing strategy to break into new markets around the world. Based on an article from Business Insider, Netflix had about 158 million subscribers worldwide with 60 million from the US and almost 98 million internationally. Netflix's original subscriber base was based solely in the United States following its IPO. A large part of its success was due to the decision to expand to international markets. The popular market prioritizes what content the company will release. In this case, we can see that a good amount of international movies and TV shows were added over the years as part of Netflix's global expansion.

## References-

- [GeeksforGeeks](#)
- [Stack\\_overflow](#)
- [Stat\\_quest](#)

