

# Capstone Project Submission

## Topic: Netflix Movies and TV Shows Clustering

### Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Name: Amol Kale Email: <a href="mailto:amolkaleak01@gmail.com">amolkaleak01@gmail.com</a> Contribution: Netflix Movies and TV Shows Clustering whole capstone project done individually.
Please paste the GitHub Repo link.
GitHub Link:- <a href="https://github.com/Amolkale01/NETFLIX-MOVIES-AND-TV-SHOWS-CLUSTERING">https://github.com/Amolkale01/NETFLIX-MOVIES-AND-TV-SHOWS-CLUSTERING</a>  Drive Link:- <a href="https://drive.google.com/drive/u/0/folders/1CLtwCuz31pqL12hIr6Q-HAXjNfIQaoP_s">https://drive.google.com/drive/u/0/folders/1CLtwCuz31pqL12hIr6Q-HAXjNfIQaoP_s</a>
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

## Summary

### Netflix Movies and TV Shows Clustering

Netflix is a subscription-based streaming service that allows our members to watch TV shows and movies without commercials on an internet-connected device. One can also download TV shows and movies to your iOS, Android, or Windows 10 device and watch without an internet connection. Netflix content varies by region and may change over time. One can watch from a wide variety of award-winning Netflix Originals, TV shows, movies, documentaries, and more. Netflix have data which were released in the past also. Not all the movies and TV shows are released first on Netflix itself

The main objective of this project is clustering of Netflix movies and TV shows in optimum number of clusters.

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. Movies and TV shows can be categorized on several factors such as genres, cast and director of movies or TV shows etc.

Steps involved:

- Data Cleaning
- Data Preprocessing
- Exploratory Data Analysis
- Data Preprocessing for Clustering
- K-Means Clustering
- Recommender System

Following conclusions can be made from project:

- Movies uploaded on Netflix are more than twice the TV Shows uploaded.
- TV shows and movies are increasing continuously but in 2019 there is drop in number of movies.
- From October to January, maximum number of movies and TV shows were added.
- Maximum number of movies and TV shows were either on start of the month or mid of the month.
- United State tops in the list of maximum number of movies and TV shows followed by India, UK and Japan.
- Maximum of the movies as well as TV shows are for matures only.
- Anupam Kher top from the list of casts having maximum number of movies and TV shows.
- Majority of movies have running time of between 50 to 150 min.
- Almost 68% of TV shows consist of single season only.
- Top 3 genres are exactly same for movies and TV shows.
- Dramas genres hit all over the world.
- 30% movies and 50% TV shows are Netflix Originals.
- Clustering done by K-Means Clustering, found optimal number of clusters equal to 9 with highest Silhouette Score.
- Recommender system using cosine similarity performs well on data.