

CapstoneProject Submission

Clustering Model

DONE BY:

Praveen. S

(Cohort Everest)

Email id: spr87654@gmail.com

Github link: <https://github.com/prav87654/Clustering-Model--Netflix-TV-shows-and-Movies>

Google Drive:

[Drive:https://drive.google.com/drive/folders/1IR4d1jbFrNhudNi2eloUdunOn046Yb1b](https://drive.google.com/drive/folders/1IR4d1jbFrNhudNi2eloUdunOn046Yb1b)

SUMMARY

Netflix, Inc. is an American subscription streaming service and production company. Launched on August 29, 1997, it offers a film and television series library through distribution deals as well as its own productions, known as Netflix Originals. As of March 31, 2022, Netflix had over 221.6 million subscribers worldwide.

Our clustering model can help us understand what number of and type of contents available in this streaming platform and its success behind by identifying the number of clusters of contents on the basis of its parameters.

This dataset consists of Tv shows and movies available on Netflix as of 2019. The dataset is collected from Flixable 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.

This dataset consists of about **7500+** records and **12** attributes

I had done the EDA and Data Visualization with the remaining features and found the independent features are not correlated with the target variable. Then Performed Outlier Detection using quantile range and also handled punctuation and stop words using nltk.

Before selecting the best models, I had done elbow method to find the optimal number of clusters which is also validated using silhouette score method.

Then I had done K-means clustering as well as hierarchical clustering with our optimal no. Of clusters **6** in which also found using the dendrogram and had displayed the cluster's visualization.