**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Summary Of NetFlix Movies and TV Shows Clustering Capstone Project** |
| --- |
| * **Initial Libraries and Dataset Import setup.** * **Data Exploration and Cleaning**   Pie chart of type(movie,TV Shows)  CountPlot of Rating  Countplot for releases over years  Top 10 genre of Movies and TV Shows on Netflix  Top 20 Actors with most number of movies/shows  Content produced Country wise  Check Increasingly Focus on Movies or Shows in recent year   * **Textual Data Processing**   Remove Punctuation  Lower Case  Remove Stopwords  Stemming  Tf-Idf Vectorizer     * **Fit the model**   KMeans  Silhouette Score  Elbow Method  DBSCAN  Hierarchical Clustering  Agglomerative Clustering  Now Model is ready for Recommendation |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Problem Statement & Approach:**  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**.**  **Conclusion**:   1. The dataset contains 7787 rows and 12 columns, cast and director columns have a lot of missing values so we dropped them and we have 10 features for the further analysis. 2. We have two types of content movies and TV shows. 3. Netflix has 69% of its content as movies, so we can say that movies are clearly more popular on Netflix than TV shows. 4. For a mature audience, there is much more movie content than TV shows. However, for the younger audience (under the age of 17),there are more TV shows than movies. 5. Netflix has started adding content since 2014,the highest number of movies and tv shows added in the year 2019,there is consistent content addition to netflix across the year. 6. The average duration of a movie on netflix is 90 minutes. 7. With respect to available content,the United States is on the top.India is at second followed by the UK and Canada. China is not even close to the top. 8. In terms of genres, Dramas is on the top followed by Comedies and Documentaries. 9. Number of movies added to Netflix is higher than that of TV shows. In 2019, netflix added 1497 movies and 656 TV shows. So there we cannot conclude that Netflix has switched focus from movies to TV shows. 10. Principal component analysis was performed inorder to reduce the higher dimensionality which improved the silhouette coefficient to 0.34118. 11. Clusters are identified for each of the records in the dataset. 12. Recommendation based on cosine similarity is done.   **Below is a summary of the model performances:**  Principal component analysis was performed inorder to reduce the higher dimensionality which improved the silhouette coefficient to 0.34118. |
| [**https://github.com/JyotiSingh0996/Netflix-movies-TV-shows-clustering-capstone\_project-4.git**](https://github.com/JyotiSingh0996/Netflix-movies-TV-shows-clustering-capstone_project-4.git) |
|  |
|  |