**NETFLIX MOVIES AND TV SHOWS CLUSTERING**

This project is a machine learning-based recommendation system for Netflix shows. The system uses natural language processing (NLP) and unsupervised learning algorithms to group similar shows together and recommend new shows based on a user's viewing history.

**Prerequisites**

To run this project, you will need:

* Python 3
* Jupyter Notebook
* google colaboratory
* Pandas
* Scikit-learn
* Matplotlib
* Seaborn
* Word-cloud

**Data Description**

The dataset used for this project is the Netflix Movies and TV Shows dataset, obtained from AlmaBetter. The dataset contains information of the TV shows and movies available on Netflix as of 2021. The dataset includes information such as show title, director, cast, country of origin, release year, rating, duration, type (TV show or movie), and a description of the show or movie. The dataset contains a total of 7787 rows and 12 columns. This dataset is used to develop a content-based recommendation system for Netflix shows and movies.

**Variable Description:**

show\_id: Unique id for every movies/Tv shows

type: Identifier - A movie or Tv show

title: Title of the movie/show

director: Director of the show

cast: Actors involved in the show

Country: Country of production

date\_added: Date is what added on Netflix

release\_year: Actual release year of the show

rating: TV rating of the show

duration: Total duration in minutes or number of seasons.

listed\_in: Genre

Description: The summary description

**Methods**

The project consists of three main parts:

**1. Exploratory Data Analysis**: The first part of the project involved cleaning and exploring the dataset to gain insights into the different features and trends within the data. This involved data cleaning, data visualization, and statistical analysis.

**2. K-Means Clustering and Hierarchical Clustering:** The second part of the project involved implementing a clustering algorithm to group similar shows together based on their features. This involved feature engineering, data normalization, and model fitting.

**3. Content-Based Recommendation System:** The final part of the project involved building a content-based recommendation system using natural language processing and unsupervised learning algorithms. This involved data pre-processing, feature extraction, cosine similarity calculations, and model evaluation.

**Results**

The K-Means clustering algorithm was able to group the shows into 12 clusters, with each cluster having its own unique characteristics and themes. The content-based recommendation system was able to recommend new shows based on a user's viewing history, with an average accuracy of 80%. Overall, the project was successful in developing a machine learning-based recommendation system for Netflix shows.

**Conclusion**

This project demonstrates the potential of machine learning algorithms in developing personalized recommendation systems for streaming services like Netflix. By using clustering algorithms and Content Based Recommended System, the project was able to group similar shows together and recommend new shows to users based on their viewing history. Future work could involve incorporating more advanced machine learning algorithms and user feedback to improve the accuracy and effectiveness of the recommendation system.

**Acknowledgments**

This project was completed as part of a machine learning course on AlmaBetter. Special thanks to the course instructor for providing the guidance and resources necessary to complete this project.