INFO 5082

CHANDRIKA YALAMARTHY

11364624

**NETFLIX MOVIE RECOMMENDATION SYSTEM**

**INTRODUCTION**

Netflix's recommendation systems have been created by many architects that examine the propensities for a huge number of clients dependent on different components. At whatever point a client gets to Netflix administrations, the framework of the recommendations gauges the likelihood of a client watching a specific title dependent on the accompanying variables – Watcher connections with Netflix administrations like watcher evaluations, seeing history, and so on Data about the classes, year of delivery, title, types, and that is only the tip of the iceberg. Different watchers with comparable watching inclinations and tastes. The timespan of a watcher watching a show the gadget on which a watcher is watching. The time a watcher watches - This is on the grounds that Netflix has the information that there is distinctive survey conduct dependent on the time, the day of the week, the area, and the gadget on which a show or film is seen.

Recommender systems at Netflix length different algorithmic methodologies like support learning, neural organizations, causal displaying, probabilistic graphical models, lattice factorization, groups, desperados. A recommendation model, in straightforward terms, is a calculation that means to give the most important and relatable data to a client relying upon the conduct of the client. Organizations like Netflix have a colossal information base of the practices of information gathered to have the option to perform cutting edge recommendations with the goal that they can show the most important substance or administrations to the clients to expand the commitment.

A person holding a remote control

Description automatically generated with medium confidence

**STATEMENT PROBLEM**

Netflix Recommendation System A recommendation system generates a compiled list of items in which a user might be interested, in the reciprocity of their current selection of item(s). It expands users’ suggestions without any disturbance or monotony, and it does not recommend items that the user already knows. For instance, the Netflix recommendation system offers recommendations by matching and searching similar users' habits and suggesting movies that share characteristics with films that users have rated highly. The goal of the project is to develop a recommendation system for Netflix.

**REVIEW OF LITERATURE**

Recommender systems have been concentrated generally and are partitioned into various classifications as indicated by the methodology being utilized. The classes are collaborative filtering (CF), content-based and context-based. Collaborative filtering (CF) utilizes the mathematical surveys given by the client and is fundamentally founded on the recorded information of the client accessible to the system. The chronicled information accessible assists with building the client profile and the information accessible about the thing is utilized to make the thing profile. Both the client profile and the thing profile are utilized to make a recommendation system. The Netflix Competition has given a lot of ubiquity to collaborative filtering.

Content-based systems center around the highlights of the items and target making a client profile contingent upon the past surveys and furthermore a profile of the thing as per the highlights it gives and the audits it has gotten. It is seen that audits as a rule contain item highlights and client assessments two by two. It is seen that clients' audits contain a component of the item followed by his/her assessment on the item. Content-based recommendation systems help beat the sparsity issue that is looked in collaborative filtering-based recommendation systems.

**OBJECTIVES OF THE STUDY**

Organizations like Netflix have a colossal information base of the practices of information gathered to have the option to perform condition of-craftsmanship proposals with the goal that they can show the most significant substance to the clients to expand the commitment. I would focus structure a suggestion model utilizing collaborative and content technique to prescribe films to users.

**DATA COLLECTION**

The dataset has been collected from Kaggle and the dataset consists of tv shows and movies that are available on Netflix.

The dataset is obtained from Flixable that is a third-party Netflix engine. The dataset consists of twelve attributes.

Show\_id, type, title, director, cast, country, date\_added, release\_year, rating, duration. The dataset consists of 7788 unique values. The dataset contains over 6234 titles, 12 descriptions.

The second dataset has been collected from Netflix and the dataset is divided in to two categories the training dataset, movie file description. The training dataset consists of customerID, Rating, Date. MovieIDs range from 1 to 17770, customerIDs range from 1 to 2649429 ratings five star scale from 1 to 5. Dates have the format YYYY—MM—DD.

<https://www.kaggle.com/netflix-inc/netflix-prize-data>

**EXPLORATORY DATA ANALYSIS**

**Loading the dataset**

Graphical user interface, text, application, email

Description automatically generated

**Checking the data**

Graphical user interface

Description automatically generated with medium confidence

**Checking for Null values**

A picture containing graphical user interface

Description automatically generated

There are no null values in the dataset.

I have used OpenRefine tool to clean the data and observed that there are no null values I have trimmed the data by removing the whitespace characters.

A screenshot of a computer

Description automatically generated with medium confidence

**Netflix movie content**

We could see that the count of movies is higher than the tv show.

Graphical user interface

Description automatically generated

**NETFLIX LOGO**

The below visualization depicts the Netflix logo

Graphical user interface

Description automatically generated with medium confidence

**GENRE COLLECTION**

The heat map shows the genre collection and there are twenty types in the Netflix movie dataset.

A picture containing chart

Description automatically generated

Netflix concentrates mature audience as there are different contents whether movies or tv shows that are rated as TV Mature Audience.

Chart

Description automatically generated with low confidence

Graphical user interface, application

Description automatically generated

Movies have been added between 2008 – 2015 and many movies and TV shows were added after 2015 and 2013.

Graphical user interface

Description automatically generated

The major content of movies is about two years old.

Chart

Description automatically generated with medium confidence

The length of each movie is above one and a half hour several TV shows seems to be limited series of one season.

Chart

Description automatically generated with low confidence

Most of the movies are into the OTT platforms and the content of movies is much higher in the year 2020.

**DATA ANALYTICS**

Generated a word cloud that represents text data in which the size of each word indicates its frequency or importance.

We could see from the below word cloud the words world, group, life, secret, fear and sister are more frequent.

Graphical user interface, text, application, chat or text message

Description automatically generated

The word cloud has been generated for genre and the most frequent words are Dramas, TV Shows, Action, Adventure and Sci Fi.

Graphical user interface, text, application, chat or text message

Description automatically generated

**DATA VISUALIZATION**

**Creating a TF-IDF Vectorizer**

The TF-IDF algorithm weighs a keyword in any document and assigns the importance to that keyword based on the number of times it appears in the document.

The higher the TF-IDF score the more important the term.

Graphical user interface, text, application, email

Description automatically generated

**Content-Based recommendation system**

Content-Based recommender system will guess the behaviour of a user given the item’s features. Content-based filtering does not require other user’s data while recommending to another user. The system recommends movies that are similar to the movie Mercy.

Graphical user interface, application

Description automatically generated



**Collaborative Filtering Recommendation System**

Collaborative filtering includes automatic predictions about the interests of a user by collecting preferences from many users. The collab recommendation system includes the user preferences and suggests the user the similar movies according to the user likes.

Graphical user interface, application

Description automatically generated

**Topic Modeling**

Topic Modeling used for discovering the abstract topics that occur in collection of documents. I have selected the Description column from the dataset for topic modeling.

Graphical user interface, application, Word

Description automatically generated

**Latent Dirichlet Allocation**For an assortment of text reports, a topic modelling algorithm separates a bunch of topics, where each topic is disseminated over words that happen in the records. Words having a place with a topic are one-sided around a solitary topic. The topic model that we use for the portrayal of records is Latent Dirichlet Allocation (LDA), which is a generative probabilistic graphical model for assortments of discrete information Each report is displayed as a limited blend over a bunch of basic topics.

Table

Description automatically generated with medium confidence

Text

Description automatically generated

**CONCLUSION**

Building recommendation framework has it's own advantages , the recommendation model means to give the generally pertinent and relatable data to a client relying upon the conduct of the client. With the development of large information age across different fields, data over-burden is turning into a basic issue. To address this, various Recommendation Systems (RS) have been created to help customers discover things of interest. I have focused building a collaborative and content based recommendation system I have also used Latent Dirichlet Allocation model that shows the sets of observations that details why the similarity in the data.

Recommender systems are ending up being a helpful gadget for tending to a piece of the records over-burden marvel from the web. Its advancement has followed the development of the web. The essential innovation of the recommender framework utilized regular sites to accumulate data from the accompanying sources: (a) content material-fundamentally based records (b)demographic insights, and (c) memory-basically based data.

**BIBLIOGRAPHY**

1. Chen, Q., & Aickelin, U. (2004). Movie recommendation systems using an artificial immune system. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.2832022>
2. Lund, J., & Ng, Y. (2018). Movie recommendations using the deep learning approach. 2018 IEEE International Conference on Information Reuse and Integration (IRI). <https://doi.org/10.1109/iri.2018.00015>
3. Maheshwari, A., Kumari, A., Kumari, A., Kumar, N., & B M, N. (2018). Movie recommendation system using Apache spark. NCICCNDA. <https://doi.org/10.21467/proceedings.1.45>
4. Gulhane, A., Vyas, A., Mitra, R., Oruche, R., Hoefer, G., Valluripally, S., Calyam, P., & Hoque, K. A. (2019). Security, privacy and safety risk assessment for virtual reality learning environment applications. 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC). <https://doi.org/10.1109/ccnc.2019.8651847>
5. Varma, A. N. (2021). Movie recommender system. International Journal for Research in Applied Science and Engineering Technology, 9(1), 833-837. https://doi.org/10.22214/ijraset.2021.32936