*The Development of a Unified Content Recommendation system based on Netflix, Hulu, Disney Plus, and Amazon Prime Video*

Rafeed Ullah

***Abstract***

*Creating a unified machine learning algorithm that helps users find new recommendations based on their total movie and tv watching or like history, without the borders of various platforms, would be a very useful tool in the age of streaming. These services have fantastic machine learning algorithms to recommend you that next show to watch. However, as streaming becomes more popular and various production studios partner or create with new streaming services, it becomes hard to have your recommendation software to point you to your best matches if you watch on multiple platforms. I plan to create a Machine Learning model that takes content data from Hulu, Netflix, Disney Plus, and Amazon Prime Video to help create suggestions for your next watch based on your previously watched or liked content.*

**Introduction**

During the golden age of streaming, new services were popping up and “brought more TV and film into our homes than ever before” [1]. Companies like Disney were continuously investing in their content portfolios to draw new users away from the old cable system into their more affordable, ad-free, digestible content. However, with so many new services consumers ended up getting multiple streaming services rather than just “cutting the cord”. When watching different content on different platforms, you lose one of the best features of streaming services, or at least limit it’s benefit: the content recommendation engine.

All these platforms will first ask what you are interested in and allow you to select a few shows to base its algorithm around. According to Netflix, titles that are consumed on the service then supersede the previously inputted preferences [2]. Additionally, recently consumed titles take precedence over older ones in terms of weight in the recommendation algorithm. While all 4 of these services stand to gain a lot from good content recommendation systems and have very sufficient systems, all of them continue to work to better their algorithms.

One such way of doing this is the enhanced usage of Neural Networks and the continued research in this field. From at least 2017 to 2020, papers on neural networks have led other algorithms based on frequency of publications [3]. Netflix and Amazon, both market leaders, both implement deep learning among other various algorithms to complete different tasks to optimize their model’s efficiency and increase time spent on the app therefore leading to better customer retention and stable profits [8]. This comes as streaming prices continue to rise with Netflix currently being the only profitable service on the market while others work to increase their prices to become more profitable [9].

**Methods**

The data for this project will be taken from the online data community, Kaggle. Shivam Bamsal, a user in the Kaggle community, provided 4 separate datasets with data from Amazon Prime, Netflix, Disney Plus, and Hulu [4, 5, 6, 7]. These datasets has everything someone might want to use in terms of associating different shows such as release date, title, cast, and most importantly: category tags. I plan to adjust the IDs of each piece of content to ensure no overlap then append all the data into a single sheet before adding the file to the data frame. Since all the category tags and cast tags are all split based on the delimiter “,”, I will most likely need to split the data based on that, so my algorithm can read all the various tags separately.

**Results and Analysis**

**Conclusion**

I will attempt to have a machine learning algorithm that is able to take at least 3 inputs of previously liked content and use that to create a list with a ranking of most likely to get watched next by the consumer. This service will be important as wages and other production costs increase, costs of content platforms will need to increase for them to be profitable. The only way for these customers to sustain their business is to ensure that despite cost increases or additional advertisements, customers will continue paying for the content that the services provide. The best way to do this is to make them spend more time on the platform and considering this want as a necessity that they would not want to get rid of. We can accomplish this by having a recommendation engine that helps various content platforms by creating a way that allows users to find the best content for them.

Future enhancements could be tying this to virtual assistant platforms so they can better understand users and help users find the content they would enjoy.

**Acknowledgements**

**References**

[1] Cranz, A., & Becker, D. (2022, December 14). The golden age of the streaming wars has ended. The Verge. Retrieved October 8, 2023, from <https://www.theverge.com/2022/12/14/23507793/streaming-wars-hbo-max-netflix-ads-residuals-warrior-nune>

[2] Netflix. (n.d.). How Netflix's Recommendations System Works. Netflix Help Center. Retrieved October 8, 2023, from https://help.netflix.com/en/node/100639

[3] Ko, H., Lee, S., Bocanegra, S., & Choi, A. (2022, January 3). A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. MDPI. Retrieved October 8, 2023, from https://www.mdpi.com/2079-9292/11/1/141

[4] Bansal, S. (2021, October 18). Netflix Movies and TV Shows. Kaggle. Retrieved October 8, 2023, from https://www.kaggle.com/datasets/shivamb/netflix-shows

[5] Bansal, S. (n.d.). Amazon Prime Movies and TV Shows. Kaggle. Retrieved October 8, 2023, from https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows

[6] Bansal, S. (n.d.). Disney+ Movies and TV Shows. Kaggle. Retrieved October 8, 2023, from https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows

[7] Bansal, S. (n.d.). Hulu Movies and TV Shows. Kaggle. Retrieved October 8, 2023, from https://www.kaggle.com/datasets/shivamb/hulu-movies-and-tv-shows

[8] Steck, H., Baltrunas, L., Elahi, E., Liang, D., Raimond, Y., & Basilico, J. (2021). Deep Learning for Recommender Systems: A Netflix Case Study. AI Magazine, 42(3), 7-18. https://doi.org/10.1609/aimag.v42i3.18140

[9] Canal, A. (2019, March 9). Your streaming bill is about to go up even more. yahoo!finance. Retrieved October 8, 2023, from https://finance.yahoo.com/news/your-streaming-bill-is-about-to-go-up-even-more-200938833.html?guccounter=1&guce\_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce\_referrer\_sig=AQAAAFYZ9uhxknmK6RJAOHi33nMu9PyDzuE7Vk3vnrV2rkhtDggUk-dZ-aBzdu29m5zbAurH8vAdo\_ZTQ0O80

[10] Ilmi, M. F. (2022, October 27). Hybrid and Tensorflow Recommender System. Kaggle. Retrieved October 8, 2023, from <https://www.kaggle.com/code/mfaaris/hybrid-and-tensorflow-recommender-system>

Other Resources:

<https://365datascience.com/tutorials/python-tutorials/predictive-model-python/>

<https://www.kaggle.com/code/vikassingh1996/netflix-movies-and-shows-plotly-recommender-sys/notebook>

<https://www.kaggle.com/code/mfaaris/hybrid-and-tensorflow-recommender-system>

<https://spotintelligence.com/2022/12/13/keyword-extraction/>

Activity Log

* 10/21/2023
  + Clean Dates from entire dataset
  + Explore Data and Set Datatypes of Dataframe
  + Remove Nulls
  + Install Tensorflow and update PIP just in case
  + Start some basic visualizations
  + Added automation tagging for content platform
  + Manually adjusted thousands of tags for rating
    - Debated if ratings should be simplified and merged, i.e. TV-G and G
  + DATA CLEANING RATING
    - Realized this is unrealistic probably