Roni Kaakaty

DSC630-T302

08/07/2021

Week 9 Exercise

Netflix PA Case Study

Netflix PA Movie Recommendations

Overview

Acquiring and retaining customers is the goal of every business, especially one that relies on a subscription-based model, like Netflix. In 2007, Netflix shifted to an online streaming service and away from their physical movie disc business due to how much money was lost to third party vendor issues. Their reliance on the postal service and mailing entities didn’t provide them with as much as value as a streaming service would. In today’s world, where everybody wants to be able to watch what they want instantly, it made sense to pivot to streaming. The patrons would happily pay more if it meant that they could watch what they want, when they want. The intent of every subscription service is to keep the customer coming back, so that they can be charged another month’s fee. The best way to retain customer’s business is to keep finding ways to entertain them. Netflix knew that if they were able to create an accurate movie recommendation model, they would be able to target the interests of their customers to keep them coming back.

Data Understanding

This section will focus on which units of analysis we will be looking for. The features that should be considered are as follows:

* Time of day
* Gender
* Age
* Genre
* Country of origin
* Type of entertainment (movie, documentary, show, etc.)
* Rating

Reasoning for each feature:

**Time of day**: Light-hearted features, comedies, or kids programming is more likely to be seen during the daytime hours. Typically, children don’t have their own profile setup and so it would be important to discern if an adult is watching or a child is watching under the adult’s profile.

**Gender**: Certain shows cater to certain genders more than others which could impact the algorithm.

**Age**: Older demographic has different interests than the younger demographic.

**Genre**: Different genres have varying degrees of popularity.

**Country of origin**: Certain shows are more popular in one country than another.

**Type of entertainment**: Identify links between different types of medias on Netflix (Are people who watch documentaries more likely to also enjoy Drama movies?)

**Rating**: What did the user rate the movie/show?

Data Preparation

Data preparation would involve getting the dataset prepared to be trained and tested. I wouldn’t include data in the same bucket if their country of origin differed because that would impact the model. Trends that may be happening in the United States, might not be applicable elsewhere, so it wouldn’t make sense to factor that rating into the same model. I also would group the data by age ranges. If I noticed that a certain user profile was watching children’s programming in the morning, but material for much mature audiences later in the day, I would surmise that a child and adult are sharing the same profile and will need to factor that in the data. Some issues that might impact the prediction ability of the model and skew the results is the password sharing that happens with subscription-based models. Friends or family members might be sharing accounts under the same user profile which would greatly impact the reliability of the recommendation. It’s hard to account for those individuals when we don’t have a profile setup for them.

Modeling

I would attempt to construct my model utilizing decision trees first, to see if I can get a good enough accuracy score. The reason I would lean towards using decision trees over other models is because the data is largely categorical, and the number of candidate inputs is very large. These models would be easy to interpret and helps the individuals working on the model to see what led to the prediction. If I still didn’t get a score high enough that I was happy with I would look to see if a model ensemble, such as Random Forests, would be a better model to use for this problem. I would utilize Python to run this code.

Deployment

I would deploy the decision trees and see how reliable the model was and if the model was able to successfully recommend a movie to a user, based on another user’s rating, due to how relatable they were. I would monitor the models and adjust the rules if I felt there were too many gaps and not enough information was being captured. I would then analyze the data to see what trends have developed and see if I was able to link users based on their similarities. Will a user who watches primarily documentaries be interested in strictly non-fiction entertainment and vice-versa.

Conclusions

The Netflix competition had a grand prize of $1 million dollars if the predictive analytics model was able to improve upon theirs by 10%. The reason the prize was so high is because the company understood the importance of being able to retain a customer’s interest. Ten percent might not seem like much, but it is quite significant when talking about the millions of people who currently use the streaming platform. Retaining that subscriber base for as long as possible will result in a lucrative revenue stream for an extended period of time. In an era where people want things at a moment’s notice and on demand, Netflix continuing to be ahead of the curve in terms of content that they provide and exclusivity, will allow them to keep their consumers and shareholders happy for quite some time.

References

“Data Science at Netflix - A Must Read Case Study for Aspiring Data Scientists.” *DataFlair*, 2 Apr. 2021, data-flair.training/blogs/data-science-at-netflix/.

Kasula, Chaithanya Pramodh. “Netflix Recommender System - A Big Data Case Study.” *Medium*, Towards Data Science, 28 June 2020, towardsdatascience.com/netflix-recommender-system-a-big-data-case-study-19cfa6d56ff5.

Stoll, Julia. “Netflix: Number of Subscribers Worldwide 2020.” *Statista*, 21 Apr. 2021, [www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/](http://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/)